Psychophysics and Modeling of Depth Perception

Arthur Jacobus Pieter Lugtigheid

Submitted in total fulfilment of the requirements
of the degree of Doctor of Philosophy

October 11, 2011

School of Psychology
College of Life and Environmental Sciences
University of Birmingham

Abstract

How do we know where objects are in the environment and how do we use this information to guide our actions? Recovering the three-dimensional (3D) structure of our surroundings from the two-dimensional retinal input received from the eyes is a computationally challenging task and depends on the brain processing and combining ambiguous sources of sensory information (cues) to depth. This thesis combines psychophysical and computational techniques to gain further insight into (i) which cues the brain uses for perceptual judgments of depth and motion-in-depth; and (ii) the processes underlying the combination of the information from these cues into a single percept of depth.

The first chapter deals with the question which sources of information the visual system uses to estimate the time remaining until an approaching object will hit us; a problem that is complicated by the fact that the variable of interest (time) is highly correlated to other perceptual variables that may be used (e.g. distance). Despite these high correlations we show that the visual system recovers a temporal estimate, rather than using one or more of its covariates.

In the second chapter I ask how extra-retinal signals (changes in the convergence angles of the eyes) contribute to estimates of 3D speed. Traditionally, extra-retinal signals are reputed to be a poor indicator of 3D motion. Using techniques to isolate extra-retinal signals to changes in vergence, we show that judgments of 3D speed are best explained on the basis that the visual system computes a weighted average of retinal and extra-retinal signals.

The third and fourth chapters investigate how the visual system combines binocular and monocular cues to depth in judgments of relative depth and the speed of 3D motion. In chapter three I show that differences in retinal size systematically affect the perceived disparity-defined depth between two unfamiliar targets, so that a target with a larger retinal size is seen as closer than a target with a smaller retinal size at the same disparity-defined distance. This perceptual bias increases as the retinal size ratio between the targets is increased but remains constant as the absolute sizes of the targets change concurrently while keeping the retinal size ratio constant. In addition, bias increases as the absolute distance to both targets increases. I propose that these findings can be explained on the basis that the visual system attempts to optimally combine disparity with retinal size cues (or in the case of 3D motion: changing disparity information with looming cues), but assumes that both objects are of equal size while they are not. In chapter 4 these findings are extended to 3D motion: physically larger unfamiliar targets are reported to approach faster than a smaller target moving at the same speed at the same distance. These findings cannot be explained on the basis of observers' use of a biased perceived distance, caused by differences in the retinal size (as found in chapter 3).

I conclude that, in line with contemporary theories of visual perception, the brain solves the puzzle of 3D perception by combining all available sources of visual information in an optimal manner, even though this may lead to inaccuracies in the final estimate of depth.

Table of Contents

List of Figures

List of Tables

1

General introduction

1.1 Introduction

The theory of natural selection maintains that, in order to survive, organisms must be adapted to their environment. Characteristics that enhance an animal's ability to survive, and therefore reproduce, will be passed on to future generations. One of the important adaptations is the development of visual perception: the ability to interpret and respond to visual information acquired from the environment. In humans, this information is particularly useful as it provides awareness of features and events within our surroundings - successful behaviour in a complex environment depends on generating the appropriate response to the physical source of the visual information. Thus, any organism that can perceive and interact with its environment appropriately (for example to see sources of food or to detect predators) will be more likely to survive and pass on its characteristics than will an organism that has no visual perception.

Visual perception begins with the requirement that an organism is sensitive to stimulation by light and has the capability to process and interpret this information. For example, in humans the eye consists of a complex layered structure at the back of the eyes that contains

light-sensitive cells called photoreceptors. Light reflected or emitted from objects in the environment enters the eye through the pupil and is focused by the cornea and lens to form a sharp projection on the retina. Photoreceptors then transduce the light into electrical signals and convey this information through the optic nerve to the lateral geniculate nucleus (LGN) which projects to the primary visual cortex where the input from the eyes is processed.

Many important day-to-day skills can be described as a function of vision. For example, the ability to discriminate different colours helps recognise objects such as food and it helps in determining boundaries and edges. More important, vision predominates in any accurate localisation of objects in space and it is unique in its ability to recover the three-dimensional (3D) structure of the environment and thereby estimate distance and depth, a process generally referred to as depth perception. The ability to perceive distance and depth is a vital skill: knowing the 3D geometry of our surroundings and the spatial properties of the objects it contains is an important determinant of our spatial behaviour. It allows us to work out where we and others are in the environment and how to navigate to a new location. In addition, we use visual information to perceive the dynamic properties of objects moving in depth, supporting visually-guided movement such as interception (e.g. playing tennis) and avoidance (dodging potentially lethal projectiles).

1.1.1 The problem of depth perception

In general, we experience the world 'out there' as a 3D world that contains objects, each of which has spatial properties, such as distance and direction, all of which can vary. Humans are extremely proficient at recovering these properties, with almost deceptive effortlessness. For example, we can usually pick up a coffee cup without thinking and we are usually able to avoid objects that obstruct our path or move towards us. It is tempting to speculate that the

Figure 1.1: Ambiguities in the monocular image. Clockwise from top left: Matt is supporting a broken chair in Geneva; Nick kisses the Great Sphinx of Giza; A giant man is touching the top of the Eifeltower; Ruben is carrying the Tower of Pisa on his back.

brain builds a detailed representation of the 3D world using the information delivered by the eyes. However, recovering depth from the two-dimensional retinal input from the two eyes is a computationally challenging task and presents the brain with intricate problems.

The problem is as follows. To estimate depth, the brain relies on signals whose interpretation is inherently ambiguous as the third dimension (i.e. depth) is not directly available from the retinal images. Specifically, the rules for projecting a 3D object onto a surface (e.g. the retina) are unambiguously defined by simple geometry. However, the inverse operation - the mapping from the retinal image to the three-dimensional structure of the world - is ill-posed because every two-dimensional image is consistent with an infinite number of three-dimensional scenes (the inverse-optics problem). Visual illusions, such as shown in figure 1.1, illustrate the brain's dilemma (for a review, see Gillam, 1980). Solving the puzzle of depth perception remains an open challenge, and as a result, uncovering the processes that support

recovering the 3D layout from profoundly uncertain retinal signals has been a central theme in vision science for the last 150 years.

1.1.2 Thesis aims

It is well established that when depth is specified by multiple cues, the visual system attempts to combined them into a unified percept of depth, thereby circumventing some of the ambiguity from single cues (see section 1.3). I am particularly interested in the processes underlying this combination process: how does the brain piece together information from different ambiguous cues in both static and dynamic environments (i.e. 3D motion)? A recent surge of interest in cue combination has led to a number of computational models. In short, contemporary theories of cue combination conceive the combination process broadly in terms of two stages, in which depth estimates are first derived from each cue independently, followed by a combination stage in which a weighted (optimal) combination of these estimates is calculated; the weight that is assigned in this combination stage reflects its relative reliability in the current visual context (Landy, Maloney, Johnston, & Young, 1995). From this framework, two distinct questions emerge that map directly onto the questions asked in this thesis. First, what sensory information is available to estimate distance, depth and motion in depth, and which of these sources of information are used? Second, in order to understand how the brain combines these cues, we must investigate their relative effectiveness in their visual context: how do different cues contribute to the final estimate of depth?

Now that the goals of this thesis are set, I will use the remainder of this chapter to discuss some principles of depth perception. I will start to review the cues that are at the visual system's disposal to judge depth. The aim is not to provide an exhaustive list; rather, I will only discuss the sources of information that are important to the experimental chapters in

this thesis. I will then discuss the major theories of cue combination. The chapter ends with an overview of the experimental chapters.

1.2 Taxonomy of depth cues

1.2.1 Oculomotor information

In exploring the world around us we constantly make many eye movements. These eye movements are necessitated by the structure of the human retina: only a small part of the retina, the fovea, contains a high density of photoreceptors and makes up the high resolution part of the retina that is responsible for high-acuity vision. Consequently, when we want to inspect details of the visual world they have to be projected onto the fovea. To do this, we change the orientation of our eyes (e.g. by means of a saccadic eye movement) so to bring a new portion of the environment onto our fovea.

Eye movements are made up of two distinct components (Figure 1.2). One is a lateral component called version, in which the eyes simultaneously move in the same direction, for instance from left to right. To change fixation to a new location in depth, the eyes move in opposite directions. These eye movements are called vergence eye movements. For example, when we look at a point that is close, we rotate our eyes inward towards the nose (i.e. the eyes converge). Conversely, when we look at a point farther away each eye rotates outward (i.e. the eyes diverge). Likewise, we contract the ciliary muscles (which reside around the cornea) to adjust the curvature of the lens, so that we can bring objects at a particular distance into focus. Changes in vergence and accommodation normally occur together and provide some useful information about the distance of objects. They are reviewed below.

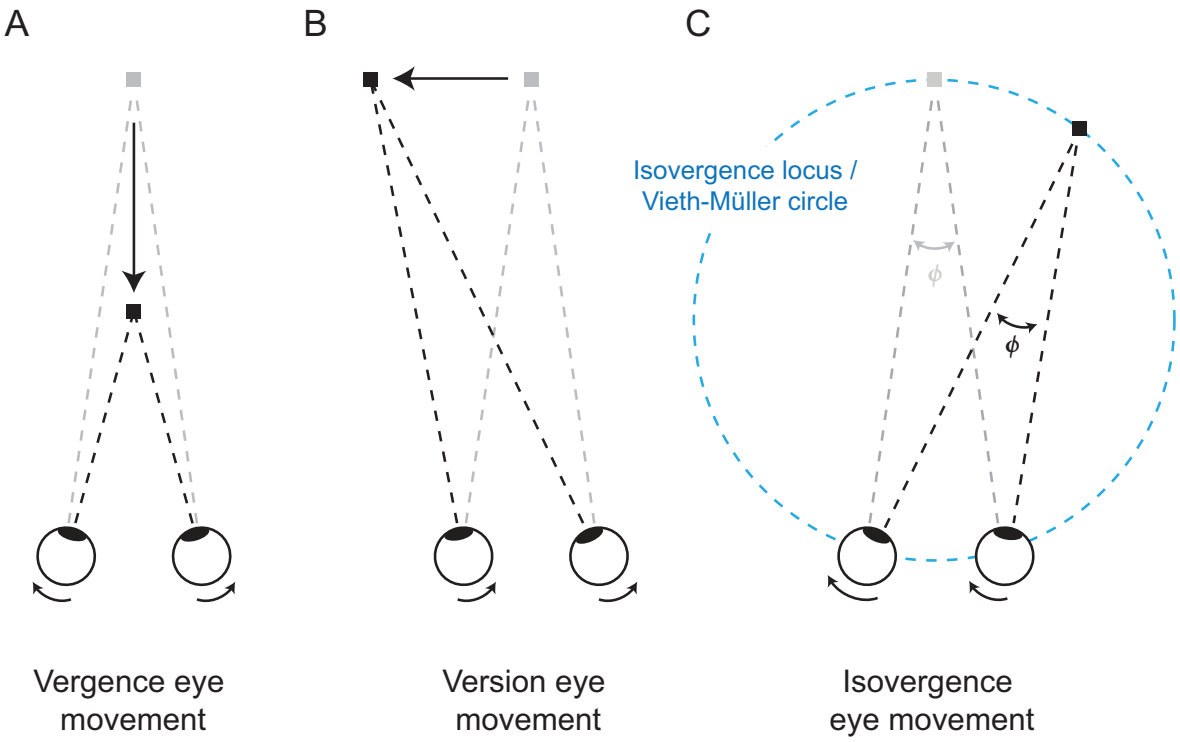

Figure 1.2: Types of eye movements. Grey squares show the initial point of fixation; black square show the new point of fixation. (A) A vergence eye movement, in which the eyes move in opposite directions to change fixation from a farther point to a closer point. (B) A vergence eye movement, in which the eyes move in the same direction to change fixation from a point straight ahead to a point to the left of the median plane through the eyes. (C) An isovergence eye movement, in which an eye movement is made that maintains a constant vergence angle. The circle that defines all isovergence points goes through the centre of the eyes and the initial point of fixation. This circle approximates the Vieth-Muller circle, which defines the theoretical horopter: the points in space which fall on corresponding points in the two retinae (the locus of isovergence passes through the centres of the eyes, the horopter passes through the nodal points of the eyes).

Convergence. We can use the orientation of the eyes somewhat like a range finder to get a rough estimate of the fixated object's distance; the point at which the ocular axes intersect specifies the fixation distance. If the inter-ocular separation and the eyes' vergence angles are known, then the distance of fixation could in principle be recovered. However, distance information from vergence is limited to a restricted range of distances, as the eye's orientation is essentially parallel at fixation distances farther than 6 meters (e.g. Foley, 1980; Collett, Schwarz, & Sobel, 1991; Tresilian & Mon-Williams, 2000; Mon-Williams & Tresilian, 1999) and differences in depth beyond this distance lead to a minimal change in vergence.

Nevertheless, in close proximity there are several ways in which the visual system could exploit the orientation to recover the distance of objects. First, the visual system might use direct extraretinal information about the vergence state of the eyes. However, although there are reports that observers make use of extraretinal information (Gogel & Tietz, 1977; Richards & Miller, 1969) judgments are often poor and there is little consistency between observers, other than a systematic tendency to underestimate distance (Gogel & Tietz, 1973; Foley, 1980). Alternatively, to judge relative distance, the visual system may use retinal information about the movement of the eyes. For example, Enright (1991, 1996) showed that observers' judgments of distance were fairly accurate when observers were made to look back and forth between two targets. Enright suggested that observers were comparing retinal disparity before and after an isovergence saccade (i.e. the vergence angles between the eyes remained the same during the saccade). In this strategy, the difference between the retinal position of the object following a saccade (the absolute disparity with respect to the fovea) is compared with where it was before the eye movement to measure disparities. Another strategy may be to use a single estimate of distance, a reference point in space, and use this to scale the differences in disparity to other objects into an estimate of depth (Foley, 1980). A third and final possibility may be that observers are sensitive to changes in vergence across saccades that have both a vergence and a version component and it may be that observers can make judgments of distance by directly measuring the change in vergence but only if they have reliable information to the orientation of the eyes before convergence changed (Brenner & van Damme, 1998).

Accommodation. Like vergence, changes in accommodation occur between near and far points; these differences in focus have been shown to convey some information about depth. For example, photographers and filmmakers often create a strong impression of depth by

simulating a blurred image of the objects that are not in the plane of fixation. Although it is clear that accommodation and image blur can potentially provide some information about distance, observers are unable to accurately judge absolute distance on the basis of accommodation in isolation (Mon-Williams & Tresilian, 1999). Other work on blur has yielded mixed results; some studies report clear contribution of blur to slant perception (e.g. Watt, Akeley, Ernst, & Banks, 2005) while others have found either no effect (Mather & Smith, 2000) or a limited effect of blur on perceived depth ordering (e.g. Palmer & Brooks, 2008).

Summarising, oculomotor depth cues are traditionally seen as unreliable cues to depth perception. Yet, a recent study - employing a Bayesian framework - showed that when blur is combined with other information (e.g. perspective cues) absolute and relative distances can be estimated fairly accurately (Held, Cooper, O'Brien, & Banks, 2010). The effectiveness of extraretinal signals to vergence is limited to a short range of distances, a limitation which can partly be overcome when other depth cues are available. Finally, stereoscopic display devices (e.g. traditional stereoscopes and 3D movies) are affected by a conflict between vergence cues and accommodation; the vergence will follow the vergence demand imposed by the stimulus, but the focus will be at the plane of the screen. This may cause distortions of depth perception and visual discomfort (e.g. Hoffman, Girshick, Akeley, & Banks, 2008) and this should be taken into account when displaying large disparities on a stereoscope.

1.2.2 Binocular information

At any given moment, with fixation on one point, we can only see part of the physical world around us. The portion that is visible to a single stationary eye is defined as the monocular visual field. This field is not symmetrical because some facial structures, such as the bridge of

the nose and the boney ridge above the eye obstruct vision in some directions. If the monocular visual fields of the two eyes are superimposed, the portion of the world that stimulates both eyes can be determined: this area is the binocular visual field. Stimulation in the binocular visual field is responsible for binocular depth perception or stereopsis. In primates and humans the eye face forward thereby maximising the binocular field. As a result of the horizontal separation between our eyes (about 6.5 cm on average), each eye registers a slightly different image of the world. The brain exploits the differences (disparities) in the retinal images to retrieve the three-dimensional layout of our environment and they are the signals that drive binocular depth perception or stereopsis (Howard & Rogers, 2002; Julesz, 1971; Wheatstone, 1852). Before combining each eye's image to form a single percept of depth the visual system must first match features in the two retinal images; it has to find the counterpart of a particular point in one image and match it to its counterpart in the other eye (i.e. the correspondence problem). If features are incorrectly matched, perceived depth does not match actual depth (i.e the "Wall paper effect").

Absolute disparities, defined for a single point, carry information about the angular position of that point relative to the centre of the fovea or point of fixation. The relative disparity between two points can be described as the difference between their respective absolute disparities, also in angular units (Figure 1.3). Our sensitivity to depth is exquisite when depth judgments are based on relative disparity as compared with when they are based on absolute disparity (Blakemore, 1970; Westheimer, 1979; Westheimer & McKee, 1978), even under impoverished viewing conditions (Julesz, 1971). An important reason for this is that changes in the vergence angles of the eyes affect a point's absolute disparity: a point that is fixated by the eyes (i.e. the retinal projection of that point falls on the centre of the fovea) has an absolute disparity of zero. However, the relative disparity between two points is unaffected by

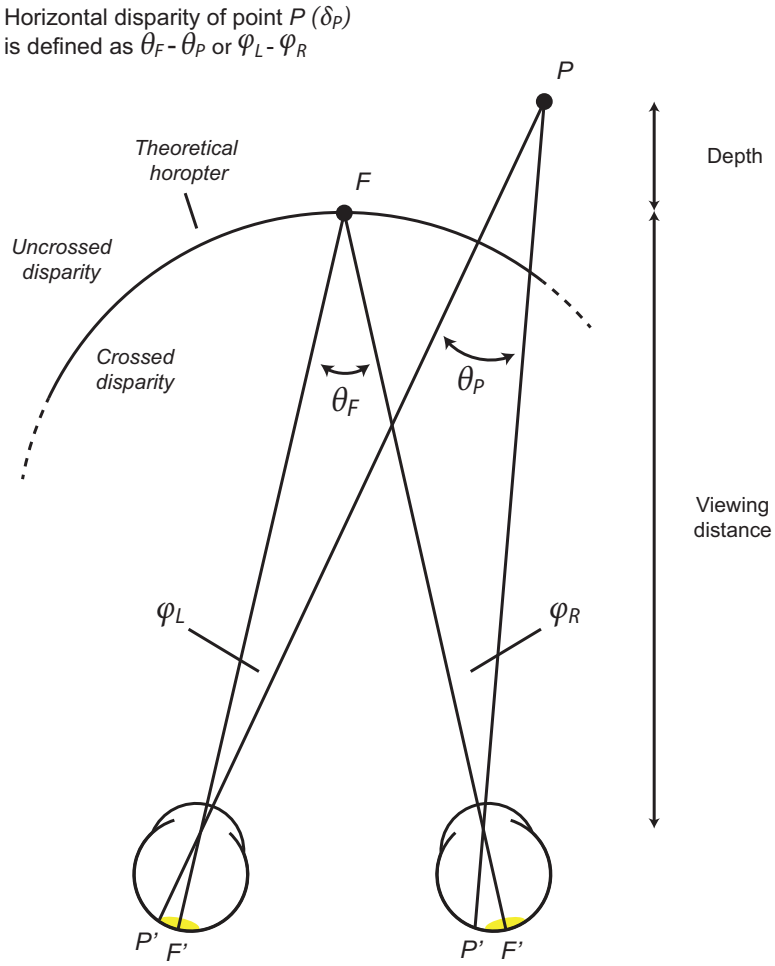

Figure 1.3: Geometry of binocular vision. The eyes fixate point F, whose image (shown as F') falls on the centre of the fovea of each eye (yellow areas). As such, its absolute disparity is zero. The retinal projections of point P (shown as P'), further away than F, fall on non-corresponding - or disparate - points in the two eyes. The relative disparity of point P is said to be uncrossed - because it lies beyond the horopter. Also see Appendix A.

changes in the orientation of the eyes. This may be one reason why the visual system exploits relative disparity to support fine depth judgments. For example, Westheimer, 1979 found that stereoacuity (the ability to discriminate depth on the basis of stereopsis) was about five times poorer when two isolated targets were presented sequentially as opposed to simultaneously; simultaneous presentation supports direct judgments on the basis of relative disparity, whereas when objects are presented sequentially the visual system has to rely on absolute disparity signals (which are affected by noise in the measurement of the orientation of the eyes).

The same conditions are true for motion-in-depth from disparity. Isolating extra-retinal signals to 3D motion, Erkelens & Collewijn, 1985b and others (Brenner, Van Den Berg, & Van Damme, 1996; Regan, Erkelens, & Collewijn, 1986; Welchman, Harris, & Brenner, 2009) have shown that changing disparity (i.e. changing the vergence demand of the entire stimu-

lus, whilst keeping its retinal size constant) only produces a sensation of 3D motion when a cue for relative disparity is available (e.g. when a static reference is presented in addition to the moving target) or when the target looms (i.e. its retinal size expands or contracts isotropically) as it moves away from or toward the observer. These and other cues to motion in depth and the visual system's use of these cues are discussed in detail in Chapters 3, 4, and 5 in this thesis. Thus, previous work has suggested that absolute disparity is not a useful cue to an object's depth. Rather, it is thought that absolute disparities are the signals that drive vergence eye movements and these disparities can correct misalignments in the orientation of the eyes by sensing the difference between the positions of the fixated point in the two retinal images (Erkelens & Collewijn, 1985b, 1985a; Masson, Busettini, & Miles, 1997; Rashbass & Westheimer, 1961; Westheimer & Mitchell, 1969).

Besides horizontal disparity, it has been shown that vertical disparities also play an important role in depth perception. Like horizontal disparity, vertical disparities result from the differential viewpoints of the left and right eyes; any point that does not lie on the median plane between the two eyes will be closer to one eye than to the other. Consequently, the retinal projection of the vertical extent presented away from this axis will be larger for the eye that is closer. These disparities increase as the eccentricity from the median plane increases. Vertical disparities can, in principle signal absolute distance (e.g. see Bishop, 1989; Mayhew & Longuet-Higgins, 1982; Longuet-Higgins, 1982; Brenner, Smeets, & Landy, 2001) but only when the stimulus is close to the observer and when the eccentricity of the stimulus is sufficiently large.

Even though it is clear that we are extremely sensitive to relative disparity information, by itself relative disparity is insufficient to specify relative distance; to correctly interpret depth from relative disparity, the measured disparity needs to be scaled by the viewing distance. Us-

ing the small angle approximation, the geometrical relationship between disparity and depth is given as follows (see Howard & Rogers, 2002):

$$\delta = \frac{I\Delta}{z^2} \tag{1.1}$$

where δ is the angular relative disparity, z is the viewing distance, Δ is the depth and I is the interocular separation. A full derivation of the equation is given in Appendix A. This equation shows that there is a non-linear relationship between disparity and depth (the so-called inverse square law), which varies with the interocular separation and the viewing distance. As a result, the same magnitude of disparity may be generated by different combinations of depth, viewing distance and interocular separations. To overcome this scale ambiguity, disparity needs to be scaled by other sources of information to the viewing distance (Foley, 1980; Ono & Comerford, 1977; Glennerster, Rogers, & Bradshaw, 1998). There are two main candidates for the source of an estimate of viewing distance: First, as noted in section 1.2.1, observers may judge distance from the vergence angle of the eyes (Brenner & van Damme, 1998; Cumming, Johnston, & Parker, 1991; Enright, 1991, 1996; Foley, 1980; Gogel & Tietz, 1977; Collett et al., 1991) and, second, the pattern of vertical disparity across the visual field (Rogers & Bradshaw, 1993; Bishop, 1989; Brenner et al., 2001; Mayhew & Longuet-Higgins, 1982). However, it should be noted that an estimate of viewing distance can also be recovered from other visual cues, including the size of familiar objects (Predebon, 1993; Sedgwick, 1986) and motion parallax (Gogel & Tietz, 1979).

In summary, the geometric relationship between depth and disparity is well understood. We know that relative disparity provides an extremely potent source of information to depth as opposed to absolute disparities which predominantly drive vergence eye movements. How-

ever, relative disparity can not unambiguously specify depth (a measured disparity is consistent with infinite combinations of absolute and relative distances) and requires scaling by the viewing distance for correct interpretation. In scaling horizontal disparity to a correct measurement of depth, both the vergence angle and a variety of visual cues contribute to the estimate of viewing distance.

1.2.3 Monocular information

Although it is clear that binocular and oculomotor cues provide strong percepts of depth, the story does not end there; we do not necessarily need two eyes to appreciate depth. If we did, people with only one eye would not be able to see depth - yet most of us maintain some appreciation of depth in when we close one eye. The reason for this is as follows: In our environment, objects are often located at different depths. When these objects are projected on the retina, these depth differences result in certain regularities in the retinal image. The visual system is sensitive to these regularities (cues) and can use them to derive an estimate of depth. The properties of these monocular depth cues have long been known by artists who use them to recreate the impression of depth on a flat canvas or screen. As a result, most paintings or films do not give us the impression of cardboard cut-outs but of a three-dimensional scene. Below is a brief overview of the most important monocular cue to this thesis, namely relative and familiar size. For a comprehensive review of other monocular cues, such as motion parallax, texture gradients, occlusion and shape-from-shading, see Cutting and Vishton (1995), Sedgwick (1986) or Howard and Rogers (2002).

Relative and familiar size. When an object's image is projected onto the retina, it's angular size (θ) depends on its distance from the eye (z)and its physical size (s):

$$tan\theta = \frac{s}{z} \tag{1.2}$$

From this equation it is clear that the retinal size of an object cannot provide any information on distance or depth, as the measured retinal size can be compatible with infinite combinations of distance and size. However, if we know the size of an object from previous experience, we can scale the retinal size into an estimate of distance. This is known as the familiar size cue (e.g. see Hershenson & Samuels, 1999; Ittleson, 1951b). When we assume that two objects are of equal size, then the ratio of their retinal sizes is directly proportional to their inverse distance ratio. This cue is known as relative size (Gogel, 1969; Ittleson, 1951b; Hochberg & McAlister, 1955; Over, 1963).

The perceived physical size of an object normally does not change with varying distance. For example, when a person walks away from us, their retinal image projection decreases in size but we do not interpret this change in size as a change in the physical size of the person (e.g. shrinking) but as a change in the distance of the person. This is because the visual system transforms retinal size into physical size, taking distance information into account. This is commonly known as size constancy (e.g. Holway & Boring, 1941) and is formalised in the size-distance invariance hypothesis, or SDIH (Kilpatrick & Ittelson, 1953). This hypothesis states that there is an approximately constant ratio between the apparent size of an object and its apparent distance in depth (Emmert, 1881; Holway & Boring, 1941; Ono, 1966). The validity of the SDIH has often been confirmed when retinal size is the only cue available. However, there are many instances in which the SDIH is invalid - for example in the case of

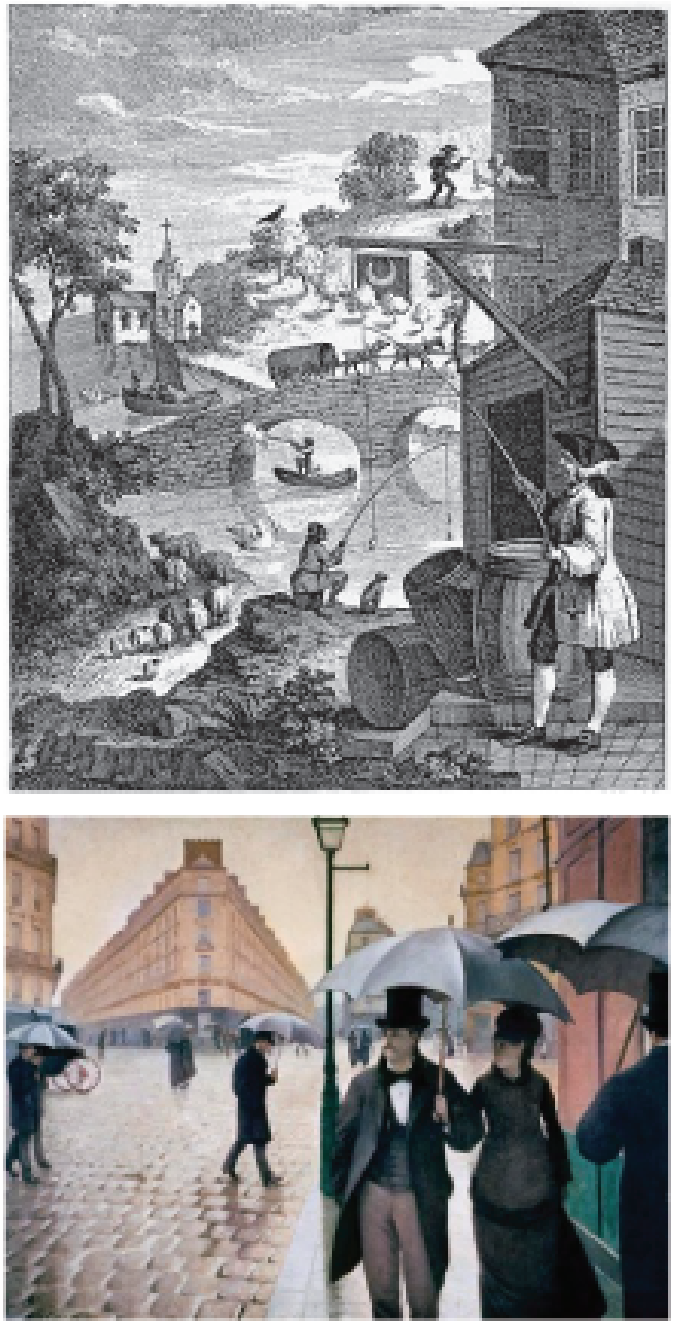

Figure 1.4: The use of monocular depth cues in art. (top) William Hogarth's "*Satire on False Perspective*" (1754). The engraving shows deliberate conflicts between cues. The subscript reads: "Whoever makes a DESIGN without the Knowledge of PERSPECTIVE will be liable to such Absurdities as are shewn in this Frontispiece". Examples of conflicts: the sign is overlapped by two trees in the background and is attached to two buildings that are not at the same distance - the man on the top of the hill is interacting with someone hanging from a window nearby and seems rather large in comparison with the church. (bottom) Gustave Caillebotte's "Paris Street, rainy day" (1877). This painting illustrates depth cues that are used correctly. For example, the texture in the stone shows a receding pattern, the umbrella occludes the lamp post and people walking in the background are depicted smaller than those on the foreground (relative size).

the moon illusion, where the moon is perceived both closer and larger when it is near the horizon than when it is at the zenith (e.g. Kaufman & Kaufman, 2000). This is commonly known as the size-distance paradox.

1.3 Depth cue combination

One recurring finding in depth perception is that judgments of depth and depth scaling become increasingly accurate as more cues are available (Bruno & Cutting, 1988; Bulthoff & Mallot, 1988; Dosher, Sperling, & Wurst, 1986; Ono & Comerford, 1977; Holway & Boring, 1941). This is due to the fact that when more than one cue is available the visual system attempts to integrate - or combine - the available information into one coherent estimate of depth. Cue combination has been studied extensively over the last few decades and has lead to a number of computational models (e.g. Bruno & Cutting, 1988; Bulthoff & Mallot, 1988; Clark, 1990; Landy et al., 1995). In general, three classes of models have been proposed. Weak fusion models compute a separate estimate of depth based on each depth cue individually on a modular basis, followed by a linear combination of the depth estimates provided by each cue. The weights assigned to each cue are proportional to each cues' reliability (Clark, 1990). Strong fusion models, on the other hand, estimate depth in a non-modular manner by combining the information from different cues in an unrestricted manner; outputs are combined without the necessity of combining the outputs of different modules (Nakayama & Shimojo, 1992).

Between these two extremes Landy et al. (1995) proposed a modified weak fusion (MWF) model. This model, the most comprehensive model to date, combines the modular aspect of weak fusion with the interactive properties of strong models. These models allow constrained interactions between cues, such as cue promotion and reweighting. For example, as discussed

previously, different cues provide qualitatively different information: vergence and vertical disparity can - under optimal conditions - provide information to absolute distance, whereas relative disparity provides only relative depth information. Due to the difference in the quality of the depth information, these cues cannot be combined; Landy et al. (1995) proposed that it is therefore necessary to transform each cue into an estimate of absolute depth. To achieve this, some depth cues must supply other cues with 'missing information'. To continue the example of relative disparity and vergence, the viewing distance signalled by vergence could be used to supply the missing information to transform disparity into absolute depth.

After the promotion stage, when all cues are made to specify absolute depth, the depth estimates from both cues can be combined. In the MWF model, the next stage is to establish the relative reliability of each cue. This may be a difficult task; in principle all depth cues are ambiguous through inherent noise in neural transmission or noise in the stimulus and may be consistent with a range of depths. Additionally, how a cue contributes to the final estimate of depth may sometimes be context-dependent. For example, it is well known that at farther distances the relative contribution of relative disparity decreases due to a less reliable signal, presumably due to the increase in variability for the estimate of viewing distance from vergence (e.g. Cutting & Vishton, 1995; Johnston, Cumming, & Landy, 1994). The reliability of cues may be established in two manners (Jacobs, 2002). First, it may be related to the ambiguity of the cue; cues that are highly ambiguous would be seen as less reliable than cues that are less ambiguous. Second, cues that are correlated to other cues (in terms of their depth estimates) may be seen as more reliable than cues that are uncorrelated.

In the final stage of cue combination, a weighted average ($\hat{S}$) of the depth estimates ($\hat{S}_i$) where the contribution of each cue (i.e. its weighting) is mediated by the reliability (w_i) of the cue. The goal of this combination process is to maximise the reliability (e.g. minimise

the variance) of the final estimate. If two cues are available (a and b), for example disparity and perspective, and provided that the noise in the individual estimates is independent and Gaussian, their combined estimate is the Maximum Likelihood Estimate (MLE):

$$\hat{S} = w_a \hat{S}_a + (1 - w_a)\hat{S}_b \tag{1.3}$$

$$\sum_i w_i = 1; \tag{1.4}$$

where the weight is proportional to the inverse variances of each cue:

$$w = \frac{1/\sigma_a^2}{1/\sigma_a^2 + 1/\sigma_b^2} \tag{1.5}$$

And the variance of the final estimate σ_{ab} is defined as the ratio of the product of each individual cue's variance to their sum, ensuring that the final variance is smaller than that of each cue:

$$\sigma_{ab} = \frac{\sigma_a \sigma_b}{\sigma_a + \sigma_b} \tag{1.6}$$

By integrating the sensory information in this manner, the combined estimate is the most reliable estimate possible (i.e. the estimate with minimal variance). As a result, this process is often referred to as 'optimal combination'. Figure 1.5 provides a visual representation of the MLE combination process of two cues.

There is an overwhelming body of evidence that sensory cues are combined optimally, for example for surface slant (Hillis, Ernst, Banks, & Landy, 2002; Hillis, Watt, Landy, & Banks, 2004; Knill & Saunders, 2003) and object shape (Ernst & Banks, 2002; Johnston, Cumming, & Parker, 1993). In addition, it has been shown that relationships between cues are not fixed and

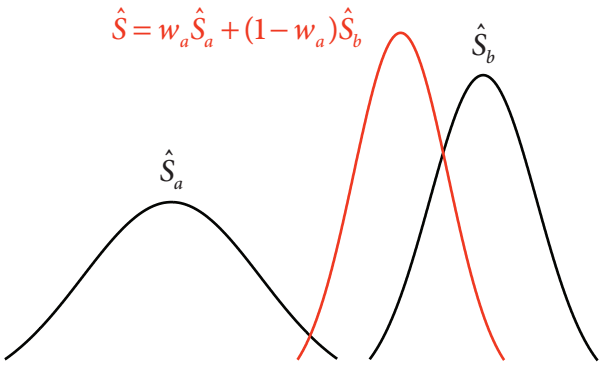

Figure 1.5: Optimal cue combination for two cues that specify different depths (i.e. there is cue conflict). Two estimates of depth ($\hat{S}$) resulting from two different sources of information (a and a) are combined. The two cues specify different depths and have difference variances. The lower variance of $\hat{S}_b$ results in a higher weighting of that cue (see equations 1.3 - 1.6) thereby 'pulling' the final estimate towards that cue. The variance of the final estimate is smaller than each individual cue's variance.

a change across the stimulus condition often results in concomitant change of cue weighting (Ernst & Banks, 2002; Hillis et al., 2002, 2004; Knill & Saunders, 2003) and when one or more cues are corrupted by the addition of noise, subjects tend to rely more in the uncorrupted cue (Alais & Burr, 2004; Ernst & Banks, 2002; Kording & Wolpert, 2004; Young, Landy, & Maloney, 1993). In conclusion, contemporary theories show that the visual system does not arbitrarily combine the depth information that is available (e.g. using a 'bag of tricks'). Rather, converging evidence confirms that sensory information is optimally combined in a manner that minimises the variance of the final depth estimate.

1.4 Overview of chapters

Chapter 2. The next chapter, the General Methods, describes the general methods and apparatus that were common to all experiments in this thesis; the details concerning each experiment are provided in their respective experimental chapters.

Chapter 3. The first experimental chapter investigates which sources of information the visual system uses to estimate the time remaining until an approaching object will hit us; a problem that is complicated by the fact that the variable of interest (time) is highly correlated with other perceptual variables that may be used (e.g. distance). Despite these high correla-

tions we show that the visual system recovers a temporal estimate, rather than using one or more of its covariates.

Chapter 4. The second experimental chapter investigates the contribution of extra-retinal signals to vergence to judgments of 3D speed. Traditionally, extra-retinal signals are reputed to be a poor indicator of 3D motion. Using techniques to isolate extra-retinal signals to changes in vergence, we show that judgments of 3D speed are best explained on the basis that the visual system computes a weighted average of retinal and extra-retinal signals.

Chapter 5-6. These chapters are closely related. *Chapter 5* investigates how the visual system combines relative disparity with retinal size. Under monocular viewing, the retinal size of an object is ambiguous with respect to its distance. Thus, differences in retinal size should not affect the perceived depth between two targets. Surprisingly, the results from this experiment show that retinal size does affect disparity-defined depth systematically, such that an object with a larger retinal size is seen as closer than an object with a smaller retinal size at the same distance. In addition, this perceptual bias increases as the ratio between the retinal sizes increases and as the absolute distance to both targets increases. The qualitative properties of these data are reasonably well described by a Bayesian cue combination model that combines relative disparity with retinal size and the vergence angles of the eyes, under the assumption that the two objects in each trial are of equal size. In *Chapter 6* these findings are extended to 3D motion: physically larger unfamiliar targets are reported to approach faster than a smaller target moving at the same speed at the same distance. These findings cannot be explained on the basis of observers' use of a biased perceived distance, caused by differences in the retinal size (as was found in Chapter 5).

Chapter 7. In the final chapter I summarise the findings in the experimental chapters and I will discuss the implications of these results in the context of the visual system's use of perceptual information and cue combination.

2

General experimental methods

Mostly the same procedures and equipment were used throughout the experiments that are presented in this thesis. The aim of this chapter is to summarise the methodology that underlies all experimental chapters. I will describe the equipment used to display and generate our stimuli. In addition, I will provide some background on the psychophysical methods that were used (and establish the terminology used in this thesis). Finally, I will provide details about our data analysis and observers. Where necessary or deviating from general methods, more details will be provided in the respective experimental chapters.

2.1 Equipment and stimulus creation

Stimuli were presented stereoscopically using a haploscope, which consisted of two 21-in. CRT displays (ViewSonic FB225f) each of which was seen in a mirror by one eye (Figure 2.1). Each mirror and CRT was mounted on a horizontal arm that rotated about a vertical axis passing through the eye's rotational centre. The face of each CRT was always perpendicular to the line of sight from the eye to the centre of the screen. Inter-pupillary spacing and vergence angle were configured for individual observers by adjusting the separation between the arms (i.e. both the mirrors and the monitors) and the rotation of the arms, respectively.

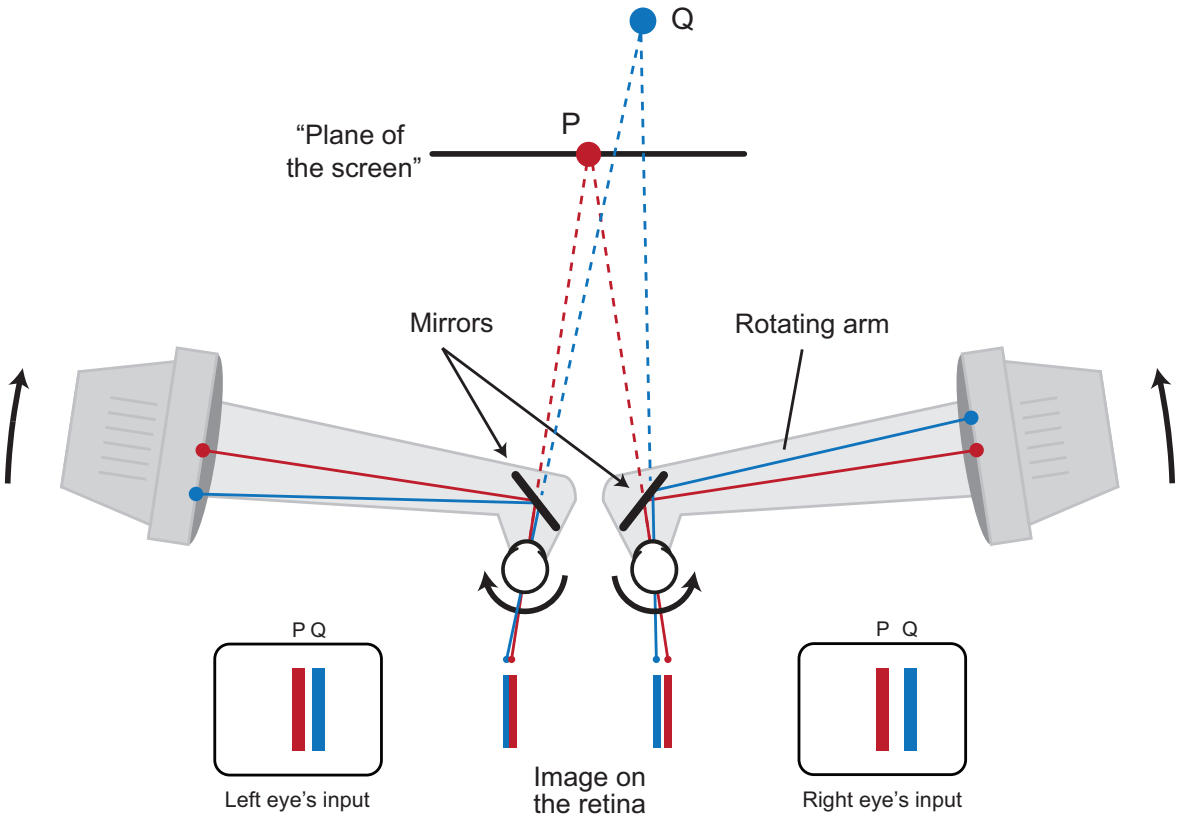

Figure 2.1: An illustration of the haploscope. Each mirror and CRT is mounted on an arm that rotated about a vertical axis through each eye's rotation centres. The face of each mirror was always at 45 degrees with respect to the CRT's. Rotation of the monitors resulted in different vergence-defined viewing distance. Here, two cylinders (P and Q) are presented stereoscopically at different virtual distances, point Q farther away than point P. The eyes fixate point P, whose image thus falls on the centre of the fovea in each eye. The insets show the input to the left and right eyes (the simulated disparity is not to scale).

Stimuli were created using OpenGL graphics libraries, implemented in the C# programming language and were rendered using anti-aliasing. The graphics card (Nvidia Quadro 4400) displayed 1600 by 1200 pixels (an individual pixel subtended about 1.5 arcmin) at a refresh rate of 100 Hz. Head movements were constrained with a chin rest to avoid information from motion parallax. Responses were recorded using a standard keyboard. Where applicable, eye movements were recorded by an Eyelink II or 1000 and were stored for offline analysis.

2.2 Psychophysical methods

The experiments described in this thesis apply traditional psychophysical techniques to complex visual stimuli. In the following section, I will provide a basic description of those used in this thesis. Visual psychophysics is the field that aims to relate (visible) physical properties of visual stimuli to the subjective psychological response that they evoke. To the psychophysicist, the brain is a "black box": the responses provided by the observer register the output of the brain, which - in principle - cannot be accessed directly. A psychophysical experiment can be seen as a sequence of stimuli-response pairs, where the stimulus is the visual display that is presented to the observer, who is then required to make a choice or judgment about the stimulus - usually through a button press. Analysing these responses, we can then link observers' subjective perceptual experience to physical or simulated stimuli.

2.2.1 The psychometric function

The fundamental building block of psychophysical methods is the psychometric function, which relates performance to the levels (often referred to as the strength or magnitude) of the stimulus parameter under investigation. To construct a psychometric function, experiments usually use the method of constant stimuli. Here, a number of varying stimulus levels are chosen that are likely to inform the experimenter of the observer's performance. These levels span a wide range, from clearly discriminable signals where observers are expected to perfectly discriminate between stimuli (either at 0% or 100%) to stimuli where observers' performance is at chance level (50%). This fixed set of stimuli is then presented multiple times (usually a minimum of 20 repetitions) in a quasi-random order that ensures that each will occur equally often. The method of constant stimuli is commonly used in combination with a two-alternative forced choice (2AFC) method. Here, observers view two stimuli (either si-

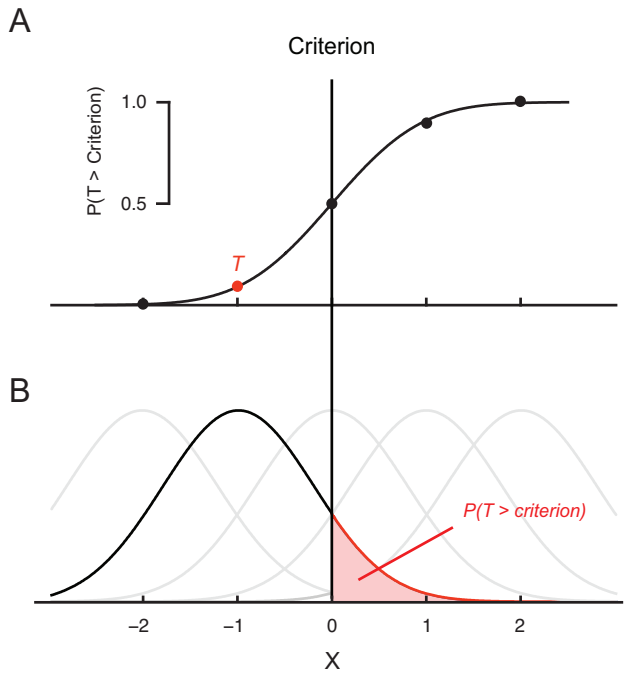

Figure 2.2: The theory underlying the psychometric function. (A) A target (T) at a stimulus value of -1 has a small chance (about .1) to be *seen* as larger than some criterion whose value is centered on zero. **(B)** The probability of T being seen as larger than the criterion is equal to the portion underneath a Gaussian distribution, centered on -1, that is larger than the criterion value. When this is calculated for all values on X it generates the psychometric function in (A). Note that a shift in the criterion results in a horizontal shift of the psychometric function - when judgments are less certain (i.e. the standard deviation of the Gaussian functions in (B) is increased) the psychometric function in (A) is less steep.

multaneously or in sequence); one is the standard (or reference) stimulus and the other is the test (or comparison) stimulus. The order in which the standard and test are presented is usually randomised. The standard is always the same in all trials but the test will differ from the standard and observers are asked to directly compare the test with the standard (e.g. "In which interval was target motion faster?"). At the end of the trial, observers are forced to choose between the two alternative choices (e.g. "First" or "Second") even when they are uncertain about their response.

Once each level has been presented multiple times, the proportion of correct responses is calculated for each stimulus level. The data are then plotted with stimulus intensity along the abscissa and percentage of correct responses along the ordinate. Note that in the case of a "subjective design" (as opposed to objective) responses are not classified in terms of percentage correct, but the percentage in which one alternative was chosen over the other (e.g. in the example given above: usually the percentage of trials in which the observer reported the comparison stimulus as faster). The resulting psychometric function is then often fitted with

a cumulative Gaussian, a curve of sigmoid shape, using two free parameters: the mean (which defines its location on the x-axis) and the standard deviation (which defines its slope).

These two parameters capture the two most basic parameters of psychophysical performance: accuracy and precision. Accuracy indicates how close an estimate is to the real presented value [1], whereas precision is related to the reliability or variance of the estimate. Precision and accuracy are often falsely used interchangeably; in fact they indicate two different sources of measurement error. The accuracy is affected by systematic error (or bias), whereas the precision is affected by random error, presumably from noise in the visual system or the visual scene. These two basic performance measures translate directly into two distinct indices that are commonly used to describe psychophysical performance: the point of subjective equality (PSE) and the discrimination threshold (also referred to as the increment threshold or the "just noticeable difference", or JND). These are described in the sections below.

2.2.2 Point of subjective equality

The *mean* of the fitted cumulative Gaussian function (the 50% point) corresponds to the PSE, which refers to the level of the test stimulus at which observers perceives it as identical to the standard. Specifically, when the test and standard are identical the observer should respond to each with equal frequency (i.e. 50% of responses). As a result, when there is a systematic bias in observers' judgments we see a shift of the location of the psychometric function on the x-axis with respect to the reference level (usually the stimulus level associated with the reference stimulus).

[1] For matching experiments, accuracy is defined with respect to a standard stimulus (either internal or external), not absolute truth. This means that if there is a bias in the 'baseline' estimate from the standard stimulus, this bias then also exists in the (matched) PSE.

2.2.3 Discrimination threshold

The *standard deviation* (or slope) of the fitted Gaussian determines the precision or discrimination threshold: the incremental change in the stimulus that produces one standard deviation change in the response (d' = 1.0). The slope therefore indicates how rapidly performance changes with changes in the stimulus strength. Discrimination threshold are the most indicative measure to observers' performance on a task. Specifically, small differences between two stimuli are often difficult to detect due to random noise. As the difference between the two stimuli increases, the probability of a successful discrimination increases until it saturates at 100% (this is why psychophysical performance is best described as a sigmoid function, such as the cumulative Gaussian). However, observers may also be very sensitive to changes in the stimulus dimension that is measured. This means that they will need smaller increments for successful discrimination between stimuli, resulting in lower threshold. A useful representation of threshold is the dimensionless weber-fraction (threshold divided by the mean) because it allows for comparison of performance across different stimulus dimensions. In the remainder of the thesis I will refer to the discrimination threshold simply as threshold. This should not be confused with an absolute threshold measure, which is an indication for performance in detection experiments in which observers detect the presence of a stimulus.

2.2.4 Choice of psychophysical procedure

Throughout this thesis we have used the method of constant stimuli to construct psychometric functions. We selected this method for its precision and reliability of its parameter estimates. Furthermore, the method of constant stimuli provides a comprehensive characterisation of psychometric performance as a function of the changes in the stimulus level, where an estimate of threshold is obtained from a fully sampled function.

2.3 Data analysis

All psychophysical data reported in this thesis has been analysed using Matlab (the MathWorks Ltd.). Psychometric functions were fitted using the psignifit toolbox version 2.5.6, which implements the maximum-likelihood method described by Wichmann and Hill (2001a). Confidence intervals for discrimination thresholds and PSE's were calculated by the percentile bootstrap method (a data resampling technique that involves a large number of repeated simulations of the experiment) implemented by psignifit, based on a minimum of 1999 simulations (Wichmann & Hill, 2001b). Where appropriate, additional advanced statistical analyses (e.g. repeated measures ANOVA's and linear regression) were conducted in SPSS. All figures were edited for publication in Adobe Illustrator.

2.4 Observer recruitment

All observers who participated in the experiments of this thesis were recruited from staff and students from the University of Birmingham and gave informed consent prior to participation. All observers were screened using a custom-programmed stereo-test to ensure that they could discriminate at least 1 arcmin of disparity in a random dot stereogram. Details about this test are provided in Appendix B. All participants received the same written instructions prior to participation.

3

Evaluating methods to measure time-to-contact [1]

Many every-day activities necessitate an estimate of the time remaining until an object will hit us: the time-to-contact (TTC). Observers' skill in estimating TTC has been studied by considering the use and combination of key visual signals (e.g. looming and disparity). However, establishing observers' proficiency in estimating TTC can be complicated, as the variable of interest (time) is typically highly correlated with other signals (e.g. target velocity or displacement). As a result, observers' responses may be based on correlates of TTC rather than on TTC itself. Here we evaluate two widely-used TTC tasks: one absolute task in which observers pressed a button to indicate the estimated TTC, and a relative task in which TTC was judged relative to a reference. We test how a wide range of experimental variables that co-vary with TTC contribute to observers' judgments. We systematically vary the correlation between TTC and its covariates and test how psychophysical judgments are affected. We show that for both absolute and relative estimation tasks, observers' responses are best explained on the basis that they judge TTC rather than one (or more) of its covariates. Our results suggest that relative tasks are preferable when assessing TTC, and we suggest a number of analyses methods to ensure that participants' judgements correspond to the variable under investigation.

[1] This chapter has been published in its current form as: Lugtigheid, A. J. & Welchman, A. E. (2011). Evaluating methods to measure time-to-contact. *Vision Research, 51*, 2234-41. Authors AL and AW conceptualised the experiment, AL was responsible for data collection and analysis. AL and AW wrote the paper.

3.1 Introduction

A key function of the visual system is to provide information about objects moving in depth so we can initiate interceptive or evasive actions (e.g. catch a ball; avoid a car crash). Frequently, the brain requires an estimate of the time remaining until an object will hit us or another object: the time-to-contact (TTC). Observers' skill in estimating this quantity has been examined by a large number of studies in both laboratory- and applied- settings. For example, applied studies have tested TTC for ball interception (e.g. Bootsma & Wieringen, 1990; Caljouw, Kamp, & Savelsbergh, 2004; Gray & Sieffert, 2005; Peper, Bootsma, Mestre, & Bakker, 1994) and the visual control of braking (e.g. Lee, 1976; Rock & Harris, 2006; Coull, Vidal, Goulon, Nazarian, & Craig, 2008), while other work has sought to isolate the key visual signals required when judging TTC (e.g. DeLucia, 1991, 2005; Gray & Regan, 1998; Heuer, 1993; Lee, Young, Reddish, Lough, & Clayton, 1983; Regan & Hamstra, 1993; Rushton & Wann, 1999; Todd, 1981).

To examine the basis of TTC judgments, observers are typically required to tune an action (e.g. a simple button press or an interceptive movement) to a visual target. However, inferring the observers' proficiency in estimating TTC in such tasks is not always straightforward, as the variable of interest (time) is typically highly correlated with other signals (e.g. the target's velocity or displacement). Thus, observer's responses may be based on correlates of TTC, rather than on TTC itself. Figure 3.1A illustrates the investigator's dilemma: varying the target's TTC (the solid diagonal) while keeping the target's starting distance (the ordinate) constant would confound TTC with the approach speed (abscissa). As a result, observers might respond on the basis of trial-by-trial variations in the target's approach speed, even though their task was to estimate TTC. A simple approach to discourage the use of covariates is to randomise the signals (e.g. speed, distance) and thereby reduce their correlation. How-

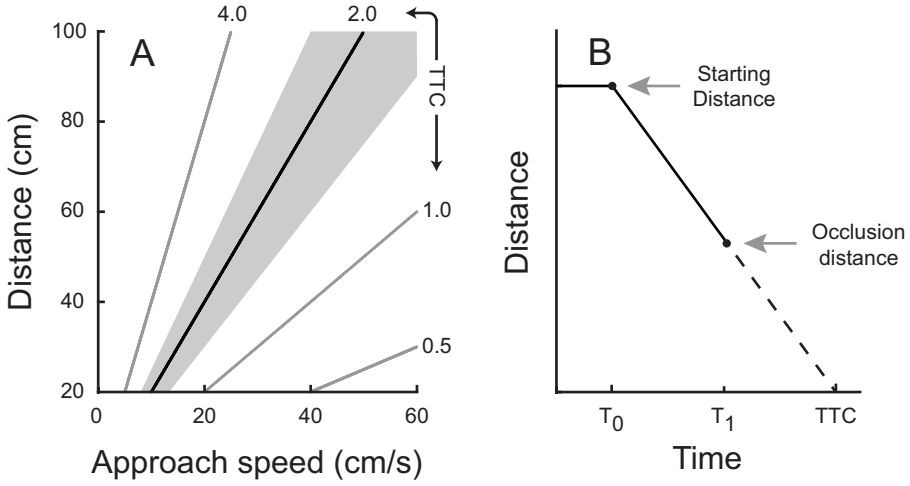

Figure 3.1: Representation of stimulus parameters involved in time-to-contact experiments. (A) TTC as a function of distance and approach speed. A single TTC can be produced from a range of combinations of distance and approach speed (solid contour lines). We sampled our values of distance and approach speed from the shaded area, with average values shown as the solid black diagonal (TTC = 2.0 s). (B) Illustration of the predictive motion paradigm (Tresilian, 1995), with distance shown as a function of time. The target remains at its starting distance for 500 ms and at time T0 starts approaching the observer. At time T1 the object is removed from the display. Had it continued along it's trajectory towards the observer (dashed line) it would hit a point between the participant's eyes at time TTC.

ever, this does not necessarily prevent observers using a covariate when responding (i.e. the lower correlation of the covariate with TTC would simply make judgments appear noisier). Therefore, it is important to test whether this manipulation is successful - evidence that many previous studies have not provided.

When presented with an approaching target, observers might exploit one or more of a range of variables to judge the likely time of impact. For instance, based on retinal size cues, they may be able to estimate TTC directly using 'tau', the ratio of the object's angular size to its rate of looming (Lee, 1976; Lee & Reddish, 1981; Lee et al., 1983; Wann, 1996; Regan & Hamstra, 1993). Alternatively, their judgments might relate to the looming rate when the approaching object is of a known size (Lopez-Moliner, Field, & Wann, 2007). Based on binocular cues, observers might use the first derivative or disparity divided by the second derivative (Regan, 2002) or the rate of change of disparity (Gray & Regan, 1998), as well as the combi-

nation of monocular and binocular signals (Gray & Regan, 1998). Given the dense intercorrelation between these signals, it can be difficult to determine whether observers' judgments relate to a full temporal estimate of TTC or rather covariates that do not unambiguously signal TTC when considered alone. One approach to the issue of covariation was developed by Regan and colleagues (Regan & Hamstra, 1993; Regan & Vincent, 1995; Gray & Regan, 1998) in which TTC was made orthogonal to other sources of information through a factorial design. For instance, Regan and Hamstra (1993) provided evidence that under monocular presentation, observers judge TTC independently from two possible covariates (retinal size and rate of expansion). While attractive, this design is unwieldy if more than two or three potential covariates are considered. Moreover, while this manipulation ensures that the looming rate is orthogonal to tau and retinal size at the start of the trajectory, this separation no longer holds as the trajectory unfolds towards the observer (the more critical period of the trial). Finally, observers in these studies were generally provided with feedback, complicating the interpretation of the results. Specifically, depending on the feedback regime, observers are able to discriminate covariates of TTC (e.g. the initial rate of expansion) with the same precision as TTC (see Regan and Hamstra (1993) Experiments 3A and 4A), making it difficult to know whether the experimental task reflects typical behaviour when judging TTC.

In this paper we seek to establish which source(s) of information participants use to judge TTC. Previous work has focused largely on the use and combination of monocular and binocular optical signals that underlie TTC judgments (i.e. looming rate, angular size and changing binocular disparity signals). Here we consider a wider range of experimental variables that also co-vary with TTC (e.g. presentation duration and occlusion distance). Our goal is to determine whether observers judge the TTC of an approaching target when instructed to do so, or rather judge one (or more) of its covariates.

In the first experiment, we use an absolute task in which observers press a button to indicate their estimate of TTC. In a second experiment, we use a relative task in which observers judge the time-to-contact relative to an auditory reference. For both experiments we consider a range of potential covariates and we systematically vary the correlation of these covariates with TTC by manipulating the amount by which covariates are randomised. We determine how performance in TTC tasks is affected by randomisation to determine whether observers' judgments rely on the actual TTC or a covariate. To preview our findings, we find that performance in both tasks suggests participants judge TTC rather than its covariates.

3.2 General methods

3.2.1 Apparatus

Stimuli were presented stereoscopically using a two-monitor haploscope in which the two eyes viewed separate 21 inch CRTs (ViewSonic FB2100x) through front-silvered mirrors. Viewing distance was 50 cm. We adjusted the haploscope so that inter-pupillary distance and vergence angle were configured correctly for each individual. Stimulus presentation was controlled by a Windows PC with an NVIDIA Quadro FX4400 graphics card. CRTs displayed 1600 x 1200 pixels at 100 Hz. Individual pixels subtended approx. 1.75 x 1.75 arcmin. The two CRTs were matched and linearised using photometric measurements. Head movements were restricted using a chin rest. Responses were collected via the PC's keyboard.

3.2.2 Stimuli

The target was a wireframe sphere (16 lines of longitude and latitude) that had a mean radius of 2 cm, randomly varied between trials from a uniform distribution in the range of ±0.2 cm (cf. Welchman, Lam, & Bulthoff, 2008). To enhance the subjective impression of 3D structure, the sphere rotated around its centre (rotation speed of 40 deg/s around the x-axis and 80 deg/s

around the y-axis). In addition to the target, a peripheral reference volume of textured cubes was visible throughout all experiments, creating the impression of viewing the target at the centre of a short tunnel. The frontal plane of the 'tunnel' was aligned with the plane of the screen and the tunnel extended 30 cm behind the screen. This provided observers with a constant stationary reference. Stimuli were created using C# and OpenGL graphics libraries and were rendered using anti-aliasing and geometric perspective projections from each eye, taking the observer's inter-pupillary distance (IPD) into account.

3.2.3 Procedure

Observers sat in the dark and viewed the motion excursion of an approaching target. At the start of each trial, the target appeared at a randomly chosen starting distance along the cyclopean line of sight. It remained at this starting distance for 500 ms to allow observers to fixate and fuse the stimulus. The target then started to approach the observer along the cyclopean line of sight at a constant (real world) speed. The target was removed from the screen at a chosen 'occlusion' distance from the observer (see Figure 3.1B). As a consequence, observers made their response based on a prediction of the target's motion (Tresilian, 1995). Observers were free to move their eyes and no feedback was provided.

3.2.4 Choice of stimulus parameters

To reduce the correlation between stimulus variables, we randomised the start distance, the occlusion distance, the TTC of the target when it started moving towards the observer and its physical size. This also randomised the approach speed of the target, the presentation duration, the rate of expansion, and the total angular expansion. To maintain a comfortable range of binocular fusion (Hoffman et al., 2008), while still allowing enough randomisation of start and occlusion distances, we set the maximum visible motion trajectory between 105

Randomisation level	Start distance (cm)	Occlusion distance (cm)	TTC (s)
0	95 ± 0.0	60 ± 0.0	2.0 ± 0.0
1	95 ± 3.0	60 ± 6.0	2.0 ± 0.3
2	95 ± 6.5	60 ± 13.0	2.0 ± 0.3
3	95 ± 8.0	60 ± 18.0	2.0 ± 0.3
4	95 ± 10.0	60 ± 20.0	2.0 ± 0.5

Table 3.1: Ranges of starting distance, occlusion distance and time-to-contact in five conditions used in Experiment 1 and 2. Values were sampled from a uniform distribution (shown here as mean ± range).

cm and 40 cm from the observer. To reduce the correlation between variables, we employed five conditions of increasing randomisation. We kept the mean value of the starting distance (95 cm), occlusion distance (60 cm) and the TTC (2 sec) constant across conditions, while systematically increasing the range of the uniform distribution from which we sampled. We included one condition in which we randomised none of the variables (Randomisation level 0, Table 3.1), one condition in which we maximised the randomisation within the range of distances we chose (Randomisation level 4, Table 3.1) and three intermediate conditions (Randomisation levels 1-3, Table 3.1). Each observer participated in each condition in a quasi-random order.

3.2.5 Variables considered as potential covariates

We considered the influence of a number of variables that could, potentially, have been used by observers when making their judgments (even though some of these potential covariates would not represent entirely rational choices). We included the spatio-temporal variables of looming rate, change in binocular disparity and the target's approach velocity. We also considered spatial variables, such as the vergence distance (expressed in angular units) and the target's retinal size, and temporal variables, such as the TTC and the presentation duration. Finally, we considered the target's total change in angular size and the total change in vergence

(i.e. the relative disparity between the starting point and the occlusion of the target). Where applicable we considered these variables both at the start of the trial (t=0) and at occlusion (t=1). We conducted repeated-measures ANOVAs in SPSS and used sphericity corrections where required. Other data processing and statistical tests were performed using Matlab (The MathWorks Inc).

3.2.6 Observers

Observers were recruited from the staff and students of the University of Birmingham (average age across all participants 26.9±4.6 years); all gave written informed consent. Observers were screened to ensure that they could discriminate at least 1 arcmin of disparity in a briefly presented (300ms) random dot stereogram.

3.3 Experiment 1: Absolute task

Perhaps the most direct measure of a person's ability to estimate TTC is to show them an approaching target for a specified time and ask them to indicate the point in time when the target would reach a specified position (e.g. hit them on the head or reach their hand). This absolute estimation approach has been taken by a number of studies (e.g. Cavallo & Laurent, 1988; Geri, Gray, & Grutzmacher, 2010; Heuer, 1993; Lopez-Moliner, Field, & Wann, 2007; McLeod & Ross, 1983; Rushton & Wann, 1999; Schiff & Detwiler, 1979), although the question of whether task-irrelevant variables (rather than TTC) were used was not addressed directly. In this experiment we use an absolute estimation task to assess observers' performance in judging time to contact. To identify the information used by observers, we vary the correlation between potentially informative variables (see 3.1) and assess how randomisation influences judgments.

3.3.1 Methods

Observers (the authors and five naive observers) viewed a single motion trajectory of an approaching target and pressed a button when they thought the target would hit them (had it continued towards them at a constant speed after being removed from the screen). Participants made 160 judgments for each of the five experimental conditions (i.e. 5 levels of randomisation). Each condition was tested in a separate experimental block.

3.3.2 Results

We compared observers' estimates of the TTC with the physically presented TTC, and examined the central tendency (the median) and spread (the median absolute deviation, or MAD) of this (typically skewed) error distribution. The median provided a measure of accuracy (i.e. how close observers' judgments were to the presented TTC) and the MAD measured precision (i.e. how reliably observers made their judgments). The most notable feature of these data was the large between-subjects variability in accuracy (3.2A): some individuals reported the target would have arrived long after it would have hit them (e.g. Observer MH responds around 800ms after it would have hit him), while others reported an arrival time before the TTC (e.g. Observer BH responds around 250ms before target arrival).

Randomising the experimental parameters did not systematically influence observers' accuracy ($F_{4,24} < 1, p = .52$; not plotted); this is expected, as the mean values remained constant so it is unlikely that a systematic bias would be introduced by our manipulation. However, parameter randomisation affected precision ($F_{4,24} = 3.246, p = .02$), with participants producing less precise responses as randomisation was increased (Figure 3.2B). This suggests that at least some information carried by the covariates contributes to participants' judgments. To determine which source(s) of information best accounted for participants' judgments, we

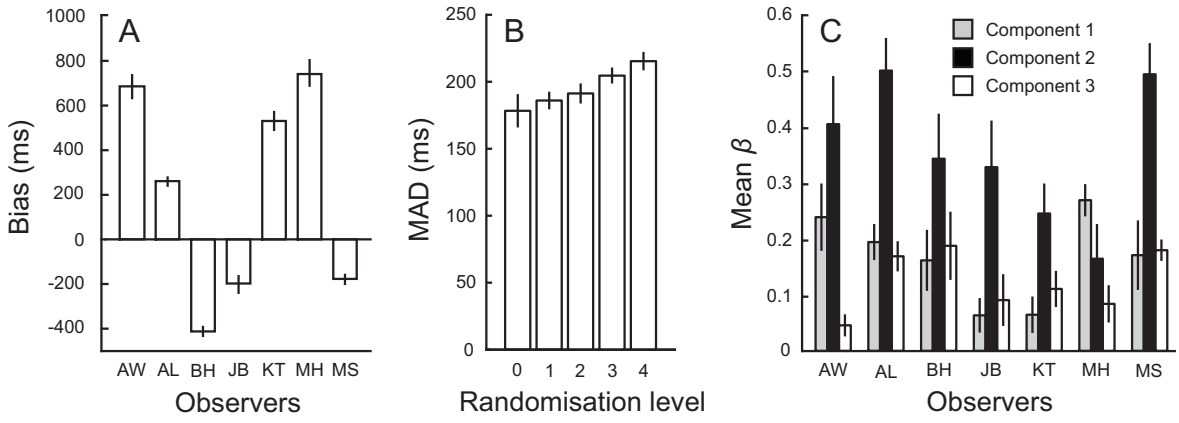

Figure 3.2: Results from the absolute task (Experiment 1). **(A)** The median error for each observer, collapsed across all levels of randomisation. Bars show bootstrapped median errors; error bars show the 95% confidence intervals. **(B)** The median absolute deviation (MAD) for the four levels of randomisation, collapsed across all observers. Bars show the bootstrapped MAD for the four levels of randomisation we used (see Table 1); error bars show 95% confidence intervals. **(C)** Mean standardised regression coefficients of the three PCA components for each observer, averaged across all degrees of randomisation. Higher values are consistent with a larger influence of the component on TTC estimates. Error bars show 95% confidence intervals for the parameter estimates.

used a regression approach. Previous studies have used stepwise multiple linear regression to determine which linear combination of variables best explains observers' responses (e.g. Gray & Regan, 1998). However, in our setting, this approach is problematic as our predictor variables are highly correlated with one another (i.e. the data have multicollinearity). As a consequence, the results of a stepwise removal or addition of predictor variables would be unstable and have poor cross-validation. To avoid this problem, we used a principal components analysis (PCA) to identify orthogonal components in the predictor variables. Having identified these components we performed a regression analysis of the data projected onto the principal component axes. Because our variables have different units, we conducted our PCA on the correlation matrix of all variables under consideration, using a varimax rotation (using Matlab's 'rotatefactors' function) to maximise the loading of each variable on one of the extracted factors while minimising the loading on all other factors. This resulted in three main components. (Note that this analysis relates to the stimulus variables generated by the

1. "Occlusion component" (about 51%)
• Looming rate at occlusion • Rate of disparity change at occlusion • Presentation time • TTC at occlusion • Vergence distance at occlusion • Retinal size at occlusion • Change in vergence (relative disparity between the start and end) • Change in angular size while the target was visible
2. "TTC component" (about 28%)
• TTC at the start of target motion • Looming rate at the start of the trial • Rate of disparity change at the start of the trial • Approach velocity of the target
3. "Start component" (about 16%)
• Vergence distance at the start of the trial • Retinal size at the start of the trial

Table 3.2: Variable composition of the extracted components ordered by the percentage variance explained (in brackets). Note, this analysis relates to values of the stimulus variables generated by the computer, not the participants' judgments.

computer, and does not yet relate to observers' judgments). The first principal component consisted of variables that mainly depended on the occlusion of the target (e.g. the looming rate at occlusion, the rate of disparity change at occlusion, the vergence distance at occlusion and the target's retinal size at occlusion); the second component consisted of the TTC, the target's approach velocity and spatio-temporal variables at the start of the trial (i.e. the initial looming rate the initial rate of disparity change); the third component consisted of spatial variables related to the start of the trial (i.e. the initial retinal size and the initial vergence distance). Table 3.2 provides an overview of the components and their variable composition.

To quantify which component best described the observers' responses we used the factor scores (i.e. the transformation of the variables into component space) of the extracted components as regressors in a multiple regression analysis, with the observers' TTC response as the dependent variable. We found that the three components accounted for observers' behaviour to a different extent ($F_{2,12} = 10.55, p < .01$), with the component relating to the TTC and the approach speed of the target (component 2) explaining most of the variance for six out of seven observers (Figure 3.2C). There was no influence of the amount of randomisation on the reliance on each component ($F_{3,18} = 2.14, p = .131$) and there was no interaction ($F_{6,36} = 3.76, p = .072$). This result suggests that participants' judgments are best explained on the basis that they use TTC, the approach speed, the initial looming rate and initial rate of disparity change rather than other covariates. In the next experiment we use a different method to examine TTC separate from all other covariates.

3.4 Experiment 2: Relative task

The results from Experiment 1 suggested large between-subjects variability in the accuracy with which individuals judge TTC in a lab-based testing situation. One means of avoiding the influence of differences in individuals' response criteria, is to ask observers to view two objects - either in sequence or simultaneously - and judge which would reach them first (e.g. DeLucia, 1991, 2005; Field & Wann, 2005; Kim & Grocki, 2006; Todd, 1981). This two-alternative forced choice format reduces the impact of an individual's decision criterion. However, this approach does not guarantee the use of TTC information: observers could make their judgments by comparing a TTC covariate (e.g. presentation duration for the two alternatives), rather by comparing the TTC variable of interest. Thus, it is important to establish which source(s) of information form the basis of observers' judgments. If a single-presentation de-

sign is used, observers can be asked to judge TTC against some internal criterion such as the mean of the stimulus set (McKee, 1981; Regan & Hamstra, 1993). Here, we take the approach developed by Gray and Regan (1998) in which observers make judgments relative to an auditory tone (also see Lopez-Moliner, Brenner, & Smeets, 2007). Having obtained TTC judgments using this task, we consider how well the different potential sources of information can account for the observers' judgments by fitting psychometric functions. We then test how psychophysical responses described in this way are affected as stimulus randomisation is varied.

3.4.1 Methods

Observers (the authors and six naive observers) viewed a single presentation of an approaching target (as in Experiment 1). In this experiment, we presented a brief auditory cue (duration of 50 ms, frequency of 1kHz) as a reference cue against which observers judged the target's time-to-contact. The timing of the reference tone could be coincident with the visually-specified TTC, or displaced from it with ± 150, 300 or 600 ms (method of constant stimuli). Observers pressed a key to indicate whether they thought the target would have hit them before or after they heard the tone (had it continued on its trajectory at a constant speed). Observers were tested in five conditions of increasing amounts of randomisation (Table 3.1).

3.4.2 Results

To assess which source of information best explained participants' judgments, we calculated psychometric functions (proportion of "after the tone" responses) expressed in terms of TTC and a range of possible covariates (cf. McKee, 1981; Harris & Watamaniuk, 1995) by binning continuous variables into equally spaced bins (e.g. Figure 3.3C). We fitted these psychometric functions with a cumulative normal (psignifit toolbox; Wichmann & Hill, 2001) and used the

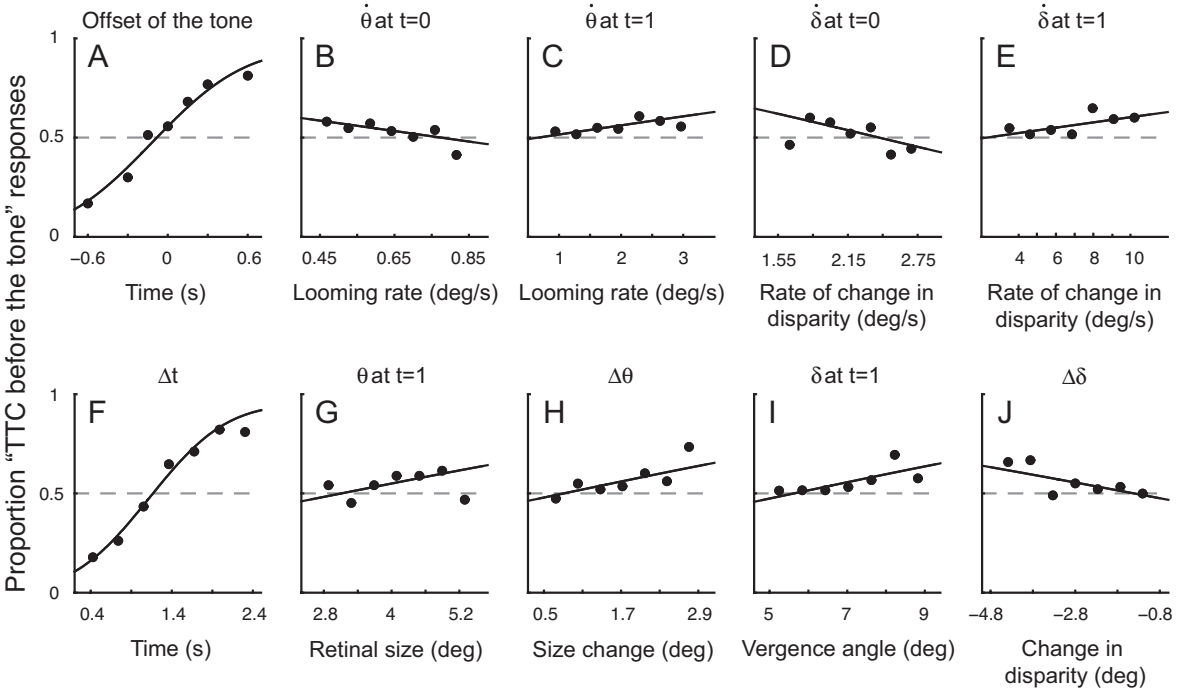

Figure 3.3: Psychometric functions from the relative task (Experiment 2), averaged across sessions for one single observer. Ten of the sources of information we considered are plotted: **(A)** the offset of the auditory cue, relative to the presented TTC; **(B)** looming rate at the start of the trial; **(C)** looming rate at occlusion; **(D)** rate of change in disparity at the start of the trial; **(E)** rate of change in disparity at occlusion; **(F)** the time interval between occlusion and the tone (ΔT); **(G)** the angular size of the target at occlusion; **(H)** the total change in the target's angular size during the visible trajectory; **(I)** the vergence angle at occlusion; and **(J)** the total change in disparity (the relative disparity between the positions at the start and at occlusion).

standard deviation parameter to quantify the discrimination threshold.

Inspecting the psychometric functions expressed in terms of different variables suggested that only two variables could reasonably account for participant's judgments: TTC and the time difference between the offset of the visual stimulus and the sounding of the tone: ΔT (Figure 3.3). To formalise this interpretation across all the participants, we calculated the 68% confidence interval for each threshold to express the range of likely underlying thresholds. We divided this confidence range by the range of stimulus values tested in the experiments. If this ratio exceeded 1, it suggested our testing range would not capture the underlying thresholds; values below 1 suggested that our data would capture thresholds reliably. For all the participants tested, the only variables that exceeded a ratio of 1 were TTC and ΔT.

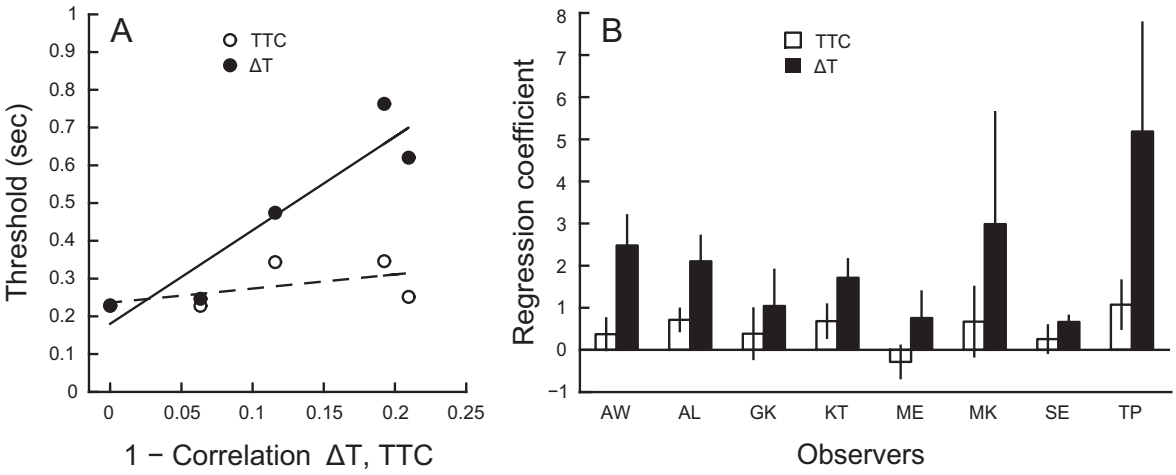

Figure 3.4: Effects of randomisation on discrimination thresholds for TTC and ΔT in Experiment 2. (A) Thresholds of TTC (open symbols) and ΔT (closed symbols) for one observer, as a function of the correlation between ΔT and TTC (here represented as 1 - correlation, such that larger values are consistent with an increase in parameter randomisation). The fitted lines represent the linear regression lines for TTC (dashed line) and ΔT (solid line). **(B)** The slope of the regression line from (A) for thresholds expressed in terms of TTC (white bars) and ΔT (black bars) for all observers. A higher value is consistent with a greater effect of randomisation. Error bars show 95% confidence intervals associated with the regression coefficients.

To investigate further which of these variables best described psychophysical performance, we examined how thresholds changed as stimulus parameters were subject to increasing amounts of randomisation. Increasing the amount of parameter randomisation had the effect of reducing the correlation between TTC and ΔT (R=1.0 with no randomisation, and R=0.78 for randomisation level 4). We found that judgments expressed in terms of TTC were relatively unaffected by variations in parameter randomisation, but this was not true when judgments were expressed in terms of ΔT (Figure 3.4). This suggests that observers' judgments are best understood in terms of judging TTC rather than simply the time interval between the disappearance of the target and the onset of the tone.

To quantify this result across observers, we fit a line to the data relating the correlation between the variables and the observer's threshold (i.e. linear regression), and then compared the slope of these lines (Fig 3.4A). A slope value of zero would indicate no influence of randomisation, while higher slope values suggest a higher influence of randomisation. For

all observers, the amount of randomisation affected thresholds related to both TTC and ΔT (Figure 3.4B). However, the influence on thresholds expressed in terms of ΔT was systematically larger ($t_7 = 3.854, p < 0.01$). This provides strong evidence that observers' responses are best expressed in terms of the TTC rather than its covariates.

In summary, asking observers to judge the TTC of an approaching target against a reference tone anchors their judgments and eliminates bias. By quantifying performance as the correlation between TTC and its covariates was reduced, we find that performance is best explained on the basis that observers judge TTC and not its covariates.

3.5 Discussion

3.5.1 Are perceptual judgments based on TTC?

In this paper we investigated which source(s) of information observers use to estimate TTC in a laboratory test. We assessed whether perceptual judgments were based on the task-relevant TTC information, or whether observers based responses on one of its many covariates. We assessed TTC judgments under two paradigms. First, we collected data using an absolute task (Experiment 1). To gain insight into the source(s) of information that best accounted for participants' judgments we conducted a principal components analysis (PCA) of the stimulus variables and regressed the resulting component scores onto the estimated TTC. For six of seven observers we found that the second component (containing the variables TTC, approach speed, initial looming rate and initial rate of disparity change) best accounted for observers' judgments. Although this provides evidence that observers responded on the basis of TTC, we could not fully dissociate TTC from other variables. As a result, it is possible that observers based their estimates on the approach speed (e.g. observers indicated a longer TTC when the approach speed was slow), the initial looming rate or the initial rate of disparity

change, rather than on the TTC.

In Experiment 2 we used a relative task and classified TTC judgments in terms of all the covariates under consideration. We compared discrimination thresholds for TTC in terms of individual covariates and found that only two variables could reasonably account for observers' judgments: the presented TTC and the time interval between target occlusion and the auditory cue (ΔT). We then showed that systematically increasing the amount of randomisation (thereby reducing the correlation between these two variables) increased discrimination thresholds for ΔT while thresholds expressed in terms of TTC were reasonably unaffected. This indicates that observers' perceptual judgments are best explained on the basis of judging TTC rather than the ΔT covariate. These results are consistent with previous reports that observers will judge TTC when asked (e.g. Gray et al., 1998). Moreover, we show that this is true for naive subjects and when no feedback is provided.

One potentially surprising result from our study is that judgments expressed in terms of looming or retinal size produce flat psychometric functions (Figure 3.3). These seems at odds with reports that humans are selectively sensitive to these cues (Regan & Hamstra, 1993). We considered the possibility that this finding may be due to our data analysis. Specifically, in Experiment 2 we measured psychometric functions for TTC judgments. To investigate the influence of other variables, we then expressed our psychometric functions in terms of potential covariates by binning the data. As a result, each bin that forms a point for the psychometric contains trials on which the auditory probe offset was at -600 ms and trials in which it was at +600 ms. At these extreme points, performance is likely at floor or ceiling: i.e. at the +600ms point observers may always respond "before", because it may be obvious -from multiple cues- that the target will arrive before the tone. Likewise, at the -600ms point observers may never respond "before". This is not a problem when we express our psychometric

function in terms of the manipulated variable, because these extreme points contribute to ceiling and floor points of the function. However, when we plot the psychometric function in terms of one of the covariates the ceiling and floor points at +600 ms and -600 ms are averaged into each data point across the range. This could potentially mask any contribution from the covariate with the result that the psychometric function would be flat.

One solution to this issue is to calculate discrimination thresholds (expressed as Weber fractions: the standard deviation of the fitted Gaussian divided by its mean) for each covariate on a reduced range of temporal offsets, excluding the data points at ±600 and ±300 ms. The standard deviation of the function fitted on this reduced range should then more accurately assess the contribution of other variables to the decision made within the critical range. Figure 3.5 shows an overview of the Weber fractions for four variables, calculated on the reduced range (blue data points) and on the full range (red data points) It is clear that Weber fractions on the reduced range are lower for most variables, thereby confirming the idea that the reduced range allows a better assessment of the contribution of covariates. However, results using this reduced range demonstrate that observers' judgments are best accounted for on the basis that they judge TTC (average Weber fraction = 0.3). Moreover, all covariates of TTC are affected by randomisation (Figure 3.4B-D): decreasing the randomisation results in a decrease of Weber fractions (this may - in part - be due to the increasing correlation of TTC with other variables) However, consistent with the results shown in Figure 3.4, the Weber fractions for TTC are relatively unaffected. This provides more evidence that observers' judgments were determined by TTC and not covariates.

As reviewed in the Introduction, recent work has suggested that looming rate is an important cue in judging interception. Here we find that looming rate does not provide a good account of our observers' judgments. This apparent discrepancy may reflect differential sam-

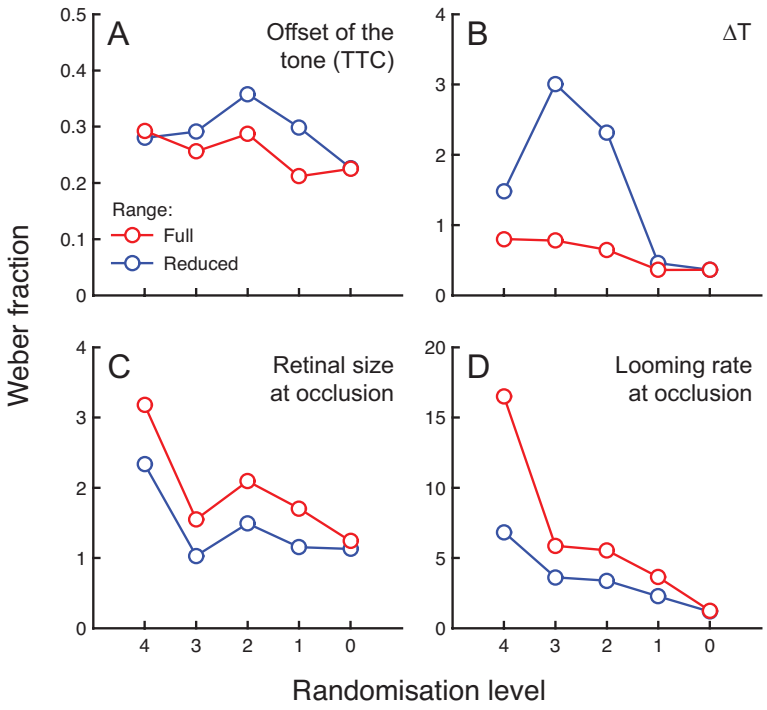

Figure 3.5: Weber fractions calculated on the full range of offsets of the auditory probe (red data points) and the reduced range excluding the extreme points (blue data points). Four of the sources of information we considered are plotted: **(A)** offset of the auditory probe, or TTC; **(B)** ΔT; **(C)** retinal size at occlusion; and **(D)** looming rate at occlusion.

pling of the approach trajectories by our study in relation to previous work. Specifically, to minimize cue conflicts, we occluded the trajectory at around 37% of its visible approach toward the observer. Had the object continued closer towards the observer, it is possible that a threshold value of looming rate may be reached that drives action initiation (Caljouw et al., 2004; Lopez-Moliner, Field, & Wann, 2007; Michaels, Zeinstra, & Oudejans, 2001).

3.5.2 How well can observers judge time-to-contact?

Having established that our observers' responses were based on TTC information, we consider their psychophysical performance. First, results from the absolute task (Experiment 1) show large systematic and esoteric errors in observers' accuracy in judging TTC using an open loop experimental task, with errors up to 850 ms (42%), with an average error of 430 ms (21%). This result is potentially alarming, considering the high precision accuracy (and precision) that is often necessitated by real-world interceptive or evasive actions. Yet, poor accuracy in estimating TTC is commonly reported in studies using absolute estimation tasks (typically 10

to 40%: Cavallo & Laurent, 1988; Schiff & Detwiler, 1979; Heuer, 1993, although see Rushton and Wann (1999) who employed fast motions in a VR setup). There are several potential explanations for these errors. First, in the context of systematic error (or bias), it is relevant to consider feedback regimes and the availability of a reference cue. Specifically, providing feedback will provide the observer with direct information about errors in their estimates and they can adjust their decision criterion to minimize their error (Karanka, Rushton, & Freeman, 2007). Similarly, a reference cue will anchor the observers' decision criteria. As we did not provide either in Experiment 1, estimates depended on individual (and unconstrained) decision criterion, resulting in large systematic and variable errors. Second, our fixed viewing distance setup involved cue conflicts (e.g. between vergence and accommodation) so estimates of approach and time-to-contact are likely to be less reliable than they would have been in a natural viewing situation.

In Experiment 2, in which observers made their perceptual judgments relative to an auditory cue, we found that errors are much lower (20-120 ms, all underestimates). Observers' judgments were reasonably precise: discrimination thresholds ranged from 180 to 600 ms; the lower range is compatible with previously reported thresholds (about 125-300 ms, Gray & Regan, 1998). The higher thresholds were associated with the naive psychophysical observers, potentially explaining the wider range of performance relative to that reported by Gray and Regan (1998) whose observers were experienced.

3.5.3 Conclusions

In summary, observers' responses under experimental tasks that use absolute estimates or relative judgments of TTC appear to be best explained on the basis that they judge the presented TTC, rather than its covariates. Our results suggest that the single-trial relative paradigm is

the more favourable method to study perceptual judgments of TTC as it reduces the impact of the systematic and variable errors that can be seen with an absolute method.

Acknowledgments

This research was supported by the Biotechnology and Biological Sciences Research Council [Grant C520620]. We thank Julie Harris for helpful comments on a previous version of this manuscript and we thank two reviewers for their careful comments on an earlier draft of the manuscript.

4

Velocity judgments of three-dimensional motion incorporate extra-retinal information [1]

When tracking an object moving in depth, the visual system should take changes of eye vergence into account to judge the object's 3D velocity correctly. Previous work has shown that extra-retinal information about changes in eye vergence is exploited when judging the sign of 3D motion. Here we ask whether extra-retinal signals also affect judgments of 3D velocity. Observers judged the velocity of a small target surrounded by a large background. To manipulate extra-retinal information, we varied the vergence demand of the entire stimulus sinusoidally over time. At different phases of vergence pursuit, we changed the disparity of the target relative to the background, leading observers to perceive approaching target motion. We determined psychometric functions for the target's approach velocity when the eyes were (1) converging, (2) diverging, (3) maximally converged and (4) maximally diverged. The target's motion was reported as faster during convergence and slower during divergence, but perceived velocity was little affected at near or far vergence positions. Thus, 3D velocity judgments are affected by extra-retinal signals about changes in eye rotation, but appear unaffected by the absolute orientation of the eyes. We develop a model that accounts for observers' judgments by taking a weighted average of the retinal and extra-retinal signals to 3D motion.

[1]This chapter has been published in Journal of Vision as: Lugtigheid, A. J., Brenner, E., & Welchman, A. E. (2011). Speed judgments of three-dimensional motion incorporate extraretinal information. *Journal of Vision*, 11, 1-11, (http://www.journalofvision.org/content/11/13/1). This chapter was a collaboration with Drs Andrew Welchman and Eli Brenner. EB and AW conceptualised the experiment. AL and EB were responsible for data collection. AL analysed the data. AL, EB and AEW wrote the paper and implemented the model.

4.1 Introduction

How do we perceive the velocity of objects moving in depth? If the eyes are fixated on a stationary target, an object's three-dimensional (3D) velocity towards and away from the observer is signalled by the change in position of the object on the left and right eyes' retinae across time. An estimate of the object's 3D velocity could therefore be derived from retinal cues, such as changes in binocular disparity (Cumming & Parker, 1994; Harris & Watamaniuk, 1995; Regan & Gray, 2009) or differences in interocular velocity (Beverley & Regan, 1973; Rokers, Cormack, & Huk, 2009; Shioiri, Saisho, & Yaguchi, 2000). Under typical viewing conditions, however, observers track the object's movement with their eyes. The resulting change in eye vergence minimizes the absolute binocular disparity of the object, reducing the magnitude of the retinal signal. Thus, it would be sensible for the visual system to combine information about retinal motion and eye vergence pursuit to estimate the object's true motion.

The visual system could derive information about vergence pursuit from the retinal slip of static scene structures, and/or from extra-retinal signals related to eye movement. However, the idea that extra-retinal signals contribute to 3D motion perception is contentious (Brenner et al., 1996; Erkelens & Collewijn, 1985b; Harris, 2006; Regan et al., 1986) and it is widely held that observers are insensitive to large changes in eye vergence when these are not accompanied by changes in relative disparity. For example, Erkelens and Collewijn (1985b) and Regan et al. (1986) reported that large changes in the absolute disparity of an extensive (30 x 30 deg) stimulus did not give rise to sensations of 3D motion, even though they induced vergence pursuit (Erkelens & Collewijn, 1985a). As a result, they concluded that extra-retinal signals about changes in eye vergence provide poor information about motion-in-depth. Further, Brenner et al. (1996) showed that observers did not perceive the 3D motion of a large object

whose absolute disparity changed by 3 deg. In these studies changing vergence signals conflicted with the absence of looming; An indication that changes in size are critical to motion in depth is that some 3D motion is perceived without retinal slip when the target does not convey looming information, for example when small targets are used (Gonzalez, Allison, Ono, & Vinnikov, 2010; Brenner et al., 1996; Regan et al., 1986; Harris, 2006; Howard, 2008). These studies suggest that judgments of 3D motion are informed by extra-retinal cues when the cue conflict is less evident.

Nefs and Harris (2008) investigated the effect of vergence pursuit eye movements on induced motion (the perception that a stationary target moves in the presence of a moving inducer). They showed that when participants pursued a fast moving inducer, induced motion of the target was tenfold higher than when they were asked to track the target. They accounted for their findings on the basis that the visual system estimates 3D motion by taking the sum of retinal and extra-retinal signals, with a gain factor attenuating the influence of the extra-retinal signal (also see Nefs & Harris, 2007). Welchman et al. (2009) showed that the retinal slip that initiated ocular pursuit is not responsible for the extra-retinal contribution to judgments of 3D motion sign (approaching or receding). Such judgments are best explained on the basis that observers combine the instantaneous retinal slip with extra-retinal (vergence) signals.

Here we extend the technique developed by (Welchman et al., 2009) to have observers make judgments about the motion of a small target that is surrounded by a large, moving background (Figure 4.1). As in the previous study, we vary the position of the background continuously over time, creating a sinusoidal vergence demand (Figure 4.2A) that induces pursuit with a high gain (Erkelens & Collewijn, 1985b). This stimulus ensures that the observers' eyes are smoothly pursuing the target (and could therefore provide extra-retinal in-

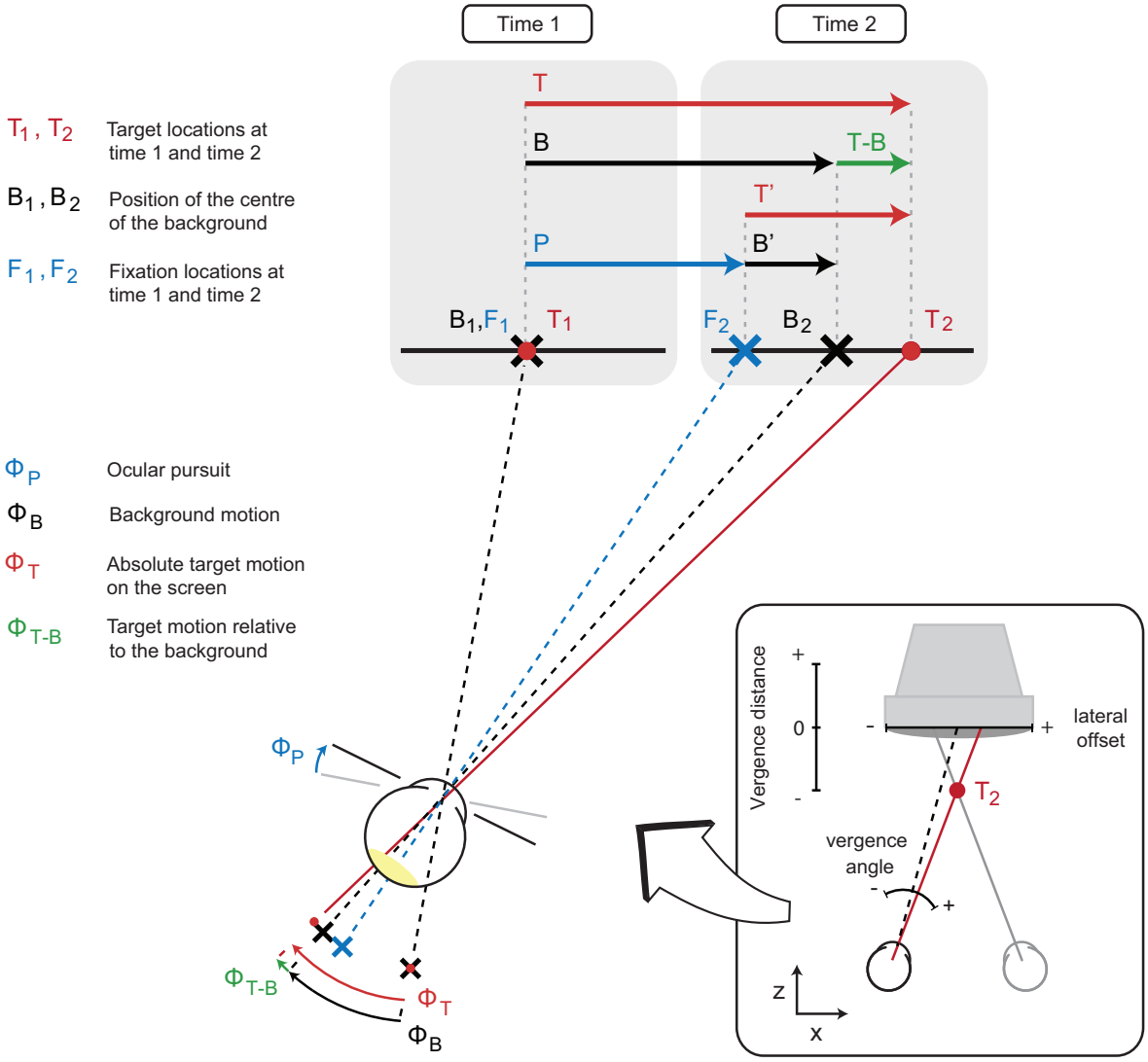

Figure 4.1: Schematic showing the lateral motion information available to the left eye for 3D motion judgments. At time 1, when the target (red dot) has not yet started to "move", the target is at the centre of the moving background (horizontal black line) and the eyes are fixated on F_1 (note that this is not necessarily representative, but is assumed for clarity). At time 2, the background has moved by extent B (black arrow) and the target by extent T (red arrow) so the target has moved relative to the background by extent T-B (green arrow). The eyes have moved by P (blue arrow) and are now fixating on F_2, slightly behind the centre of the background (B_2). This 'lag' causes a retinal flow of the background, B', and a retinal slip of the target, T'. The inset shows the interpretation of such lateral target motion in terms of motion in depth, assuming that the right eye sees a mirror symmetrical image.

formation) during the test portion of the experiment in which the target starts to move relative to the background (Figures 4.1 and 4.2B). By using a large background stimulus that has a constant retinal size, we ensure that observers cannot perceive their eye vergence pursuit due to the conflict between vergence changes and the absence of looming (Erkelens & Collewijn,

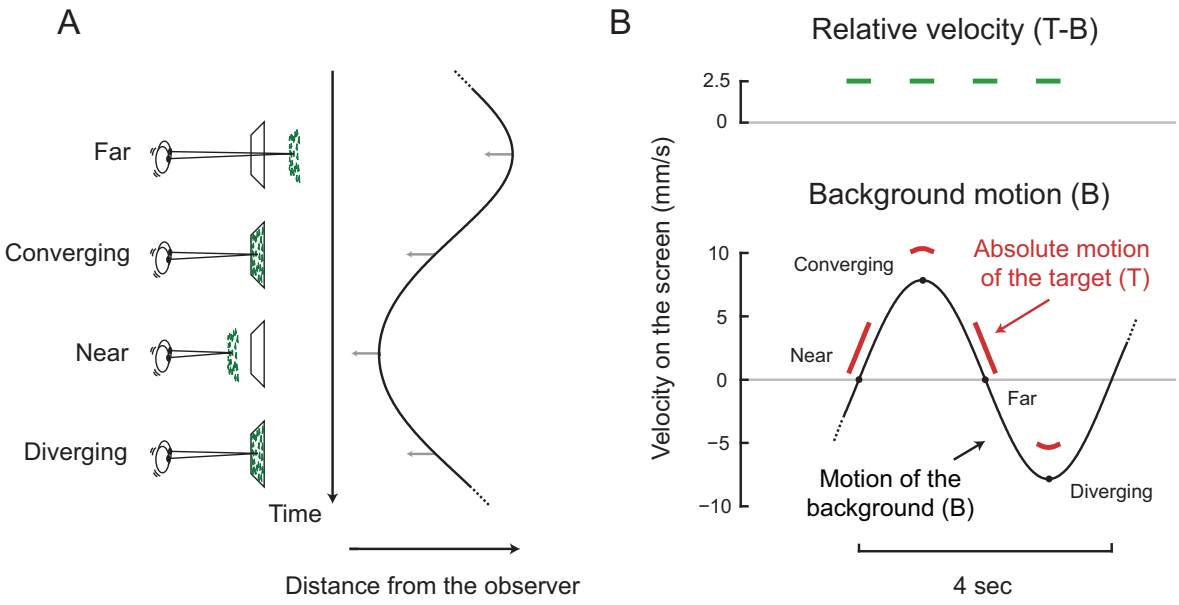

Figure 4.2: An illustration of how movement of the target and background were used to separate retinal and extra-retinal cues to motion estimation. (A) An illustration of the motion in depth of the target and background in each condition. The background moved back and forth sinusoidally, in opposite directions in the two eyes, throughout the entire experiment (frequency = 0.25Hz). In terms of binocular cues this corresponds with oscillations in depth, but due to the absence of looming these oscillations are not perceived. Target motion was perceived when we changed the relative disparity of the target with respect to the background. We did so at four phases of the background's oscillation (orange arrows): far, near, converging or diverging. (B) Velocity of the target with respect to the background (upper panel) and of the targets and the background (lower panel) for the left eye. The same relative velocity corresponds to different target velocities in the four conditions.

1985b; Regan et al., 1986; Welchman et al., 2009). Thus any influence of vergence pursuit on the interpretation of changes in relative disparity in terms of judged velocity would indicate that extra-retinal signals contribute directly to such judgments.

In our experiment, we briefly move the target with respect to the background - thereby introducing a relative retinal motion component (and thus changing relative disparity) in addition to the absolute motion of the whole stimulus (i.e. the sinusoidal displacement of the target and background). We measure judgments of approach velocity under four conditions: when the eyes are pursuing in converging (approaching) or diverging (receding) directions and when the eyes are at the maximum (near) and minimum (far) vergence excursions. Thus, we contrive that the same magnitude of the retinal cue (changing relative disparity) is com-

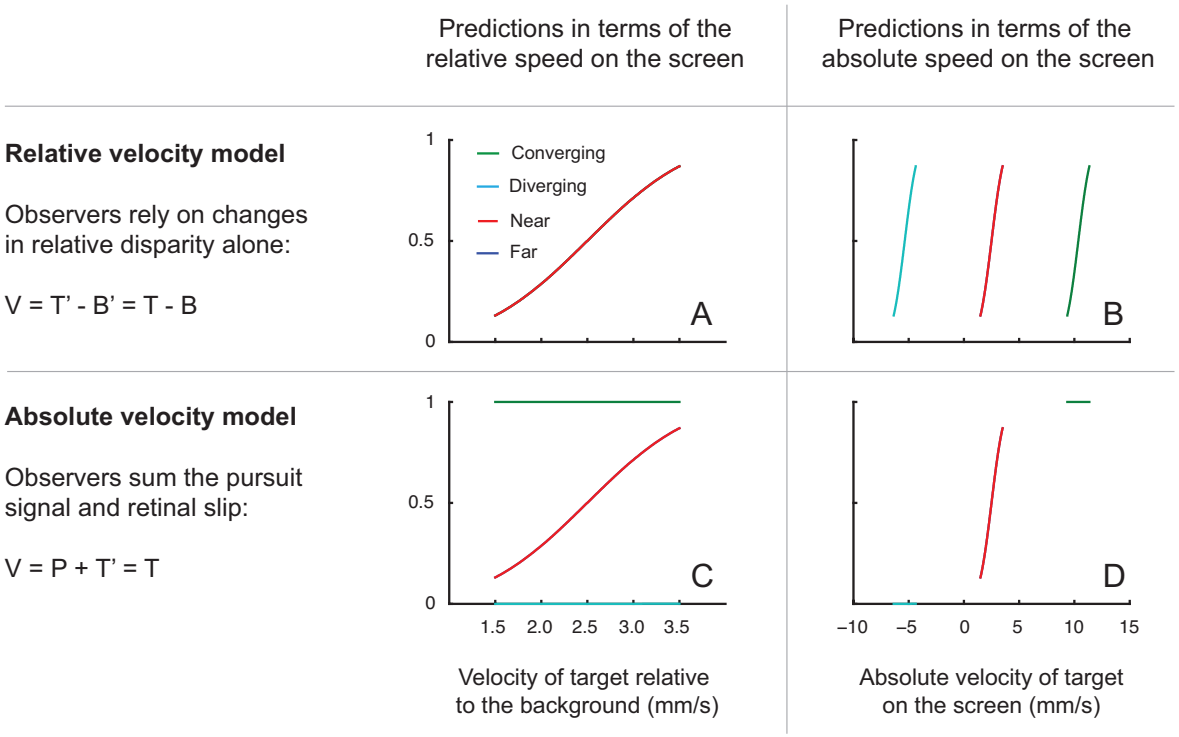

Figure 4.3: Model predictions for a relative velocity model (A,B) and an absolute velocity model (C,D). The relative velocity could be judged by taking the difference between the retinal slip velocity of the target (T') and the retinal flow velocity of the background (B'). The absolute velocity of the target could be judged by summing the pursuit velocity (P) and the retinal velocity of the target (T'). Note that since T' = T - P and B' = B - P, the predictions for the perceived velocity are T-B and T, respectively, irrespective of the pursuit velocity. Predictions are shown both in terms of the relative velocity (A,C) and in terms of the absolute velocity (B,D) on the screen. Whenever curves overlap only the red curve is visible.

bined with different magnitudes of the extra-retinal cue (Figure 4.2B, bottom panel).

Given the presence of both retinal and extra-retinal cues to 3D motion, we consider two potential models for the visual system's use of this information. First, observers might ignore all extra-retinal information, so that judgments of the target's approach velocity depend only on retinal velocity and are unaffected by differences in vergence pursuit. This relative velocity model would take as its inputs the retinal flow velocities of the background (B') and of the target (T'). Velocity judgments would be calculated as the difference between T' and B'. If so, eye pursuit velocity (P) has no bearing on the observer's judgment, because the difference between T' and B' is unaffected by adding a constant to each. This retinal model would predict psychometric functions from our four experimental conditions that lie on top of each other

when expressed in terms of relative velocity (Figure 4.3A). Alternatively, observers might judge the velocity of the target by combining the pursuit velocity (P) with the retinal velocity of the target (T'). This absolute velocity model describes the total vergence demand of the target, and predicts that judgments of velocity will be faster during convergence, and slower during divergence (see Figure 4.3D). Both of these models ignore the fact that the perceived velocity should depend on the static convergence of the eyes (i.e. the viewing distance). We will examine this scaling issue in the discussion section.

4.2 Methods

4.2.1 Observers

Two of the authors and four naive observers who were recruited from staff of the Faculty of Human Movement Sciences at the VU University of Amsterdam took part in the study. They all had normal or corrected-to-normal vision and were screened to ensure that they could discriminate 1 arcmin of disparity in a briefly (300 ms) presented random dot stereogram.

4.2.2 Apparatus

Images were presented stereoscopically on a mirror-stereoscope with two 24-inch CRT (Sony GDM-FW900) monitors, each seen by one eye through a mirror. The monitors displayed 1096 by 686 pixels at a refresh rate of 160 Hz (an individual pixel subtended about 3.1 arcmin). The distance from the observer's eyes to the monitors was about 50 cm. Observers responded by pressing keys on a keyboard. Binocular eye movements were recorded using an Eyelink II eye tracker (SR Research Ltd.) at a sampling rate of 500 Hz.

4.2.3 Stimulus

Observers fixated a small blue target dot (diameter = 7 arcmin), surrounded by a large background (20 cm / 22 deg wide, 30 cm / 31 deg high) of randomly positioned green triangles (side length 1.7 cm / 2 deg), avoiding a small region (3 cm / 3.4 deg wide, 1.5 cm / 1.7 deg high) around the target. We masked visible changes in the position of the background by rotating the triangles around their centres at a speed of 45 deg/s. Half of the triangles rotated clockwise and the other half rotated anti-clockwise. To measure the influence of extra-retinal signals the observers' eyes had to be smoothly pursuing the background in depth. We induced these vergence pursuit eye movements by continuously varying the lateral positions of the left and right eyes' images in counter-phase following a sinusoidal profile (frequency = 0.25 Hz). The amplitude of the lateral movement of the background was ±5 mm (34 arcmin), corresponding to a movement in depth from about 9 cm behind the screen to 7 cm in front of the screen. This corresponds to a vergence change of about 1.14 deg (or 0.57 deg in each eye).

To ensure that the modulations of absolute disparity in the background were imperceptible, we kept the retinal size of the background constant. This created a conflict between monocular and binocular cues for the background's simulated position in depth (looming cues signalled no 3D motion although binocular cues signalled motion). While looming information is significant for the large background, for our small target its influence is negligible (the maximum change in retinal size of the target is 2% of 7 arcmin, which is well below our rendering resolution). In previous studies, we (Welchman et al., 2009) and others (Erkelens & Collewijn, 1985b; Regan et al., 1986) have shown that keeping the retinal size of the background constant successfully prevents observers from discriminating approaching or receding motion (neither was there any perception of change in the apparent size of the stimulus). To further ensure that there was no relative retinal motion from static objects and thus that eye

movement was indicated by extra-retinal signals rather than relative retinal slip from static structures we took a number of precautions. Specifically, the experiment was conducted in full darkness and any residual light from objects within the field of view was removed by surrounding the CRTs and mirrors in dark cloth. In addition, we reduced the luminance of the CRTs to ensure that observers could not see the 'black' background illumination of the displays nor could they see the mirrors. Finally, we ensured that movement of the background stimulus did not reach the edges of the displays, ensuring there were no cues from occlusion that would arise if the stimulus moved off the screen.

4.2.4 Procedure

On each trial, the target's disparity relative to the background changed at one of five rates. The target always moved to the right across the background in the left eye and to the left across the background in the right eye, consistent with approaching 3D motion. Once the target started moving relative to the background, its colour also changed from blue to red. After the target approached the observer for 300 ms it disappeared, but the moving background remained visible. The blue target reappeared after 1 sec, moving on its sinusoidal profile with zero disparity with respect to the background. We used five rates of relative disparity change, spaced in steps of 0.5 mm/s of lateral movement around the mean rate of 2.5 mm/s. This corresponds to rates of change of disparity of about 21 to 48 arcmin/s. In the interest of consistency, we will report our results in units of mm/s on the screen, as we experimentally manipulate the changing disparity information by means of lateral motion on the screen (in opposite directions in the left and right eyes). On each trial, observers judged whether the velocity of the target was faster or slower than the mean velocity of the stimulus set (cf. McKee, 1981). We measured psychometric functions for approach velocity in four interleaved conditions: when

the eyes were moving to (a) converge or (b) diverge and when the eyes were at the end-points of their trajectories in (c) near and (d) far vergence positions (see Figure 4.2A for a cartoon).

We prevented dark-adaptation by presenting a white screen every 10 trials (approximately every 80 seconds) for 5 seconds. During this interval, we also calibrated the eye tracker by displaying an isolated black fixation target that jumped back and forth laterally every 500 ms by ±0.57 deg (the amount that the eyes were to move in opposite directions whilst pursuing the background). The median version response for each eye was considered to correspond to this distance. Each observer completed 400 trials (4 conditions, 5 stimulus levels, 20 trials) in two sessions. As it was unlikely that observers formed a reliable criterion for the mean velocity within 10 trials, we discarded the first 10 trials from each session.

4.2.5 Eye movement recording and analysis

During the experiment, observers were instructed to maintain fixation on the target. We recorded the left and right eye positions. To analyse these eye movement data, we first calibrated raw gaze positions by manually selecting fixations in calibration blocks and then converted these to degrees of visual angle. Pre-processing of eye movements involved the removal of trials in which blinks or saccades occurred during or shortly (200 ms) before or after target presentation (2% of trials) and trials in which the eye position data were excessively noisy (5% of trials) and did not resemble fixations or saccades, potentially due to instability in the eye tracker's estimate of the eye position. Eye trace signals were screened 'blind' in that neither the experimental condition nor the observer's psychophysical response was known when inspecting the eye movement traces. The removal of individual trials due to blinks and noise required the agreement of two of the authors. We calculated horizontal vergence as the right minus the left horizontal eye position, with each position being related to the positions when

fixating the screen centre (so negative values for vergence correspond to positions that are nearer than the screen).

4.3 Results

4.3.1 Eye movements

Our first analysis investigated how well our observers made vergence pursuit eye movements in response to the large moving stimulus. Figure 4.4A shows the eye position data for the critical part of a single trial in which the target was presented while the eyes where converging. To characterise the gain and the phase lag of vergence pursuit, we first combined the parts of the trajectory that we considered the critical portion of each trial (i.e. the eye trace during target presentation ± 200 ms) and then fitted a sine function to the median vergence response across observers with the function's amplitude and phase as free parameters. Vergence pursuit gain was then calculated as the peak amplitude of the fitted sine function divided by the peak amplitude of the vergence demand. Phase lag was calculated as the difference in phase between the best fitting sine and the vergence demand of the stimulus.

We found that observers made accurate vergence pursuit movements in response to the changing absolute disparity of the stimulus (Figure 4.4B), with a vergence gain of about 0.97. This is in line with previous studies that used a frequency of 0.25 Hz, which reported pursuit gains that approached unity for velocities of up to 1.5 deg/s (Erkelens et al., 1985b). We found that there was a mean delay (pursuit lag) of approximately 100 ms between the changing disparity of the background and the vergence response.

4.3.2 Perceived 3D velocity

We next examined velocity judgments under the four experimental conditions (convergence, divergence, near and far vergence). We obtained psychometric functions for approach veloc-

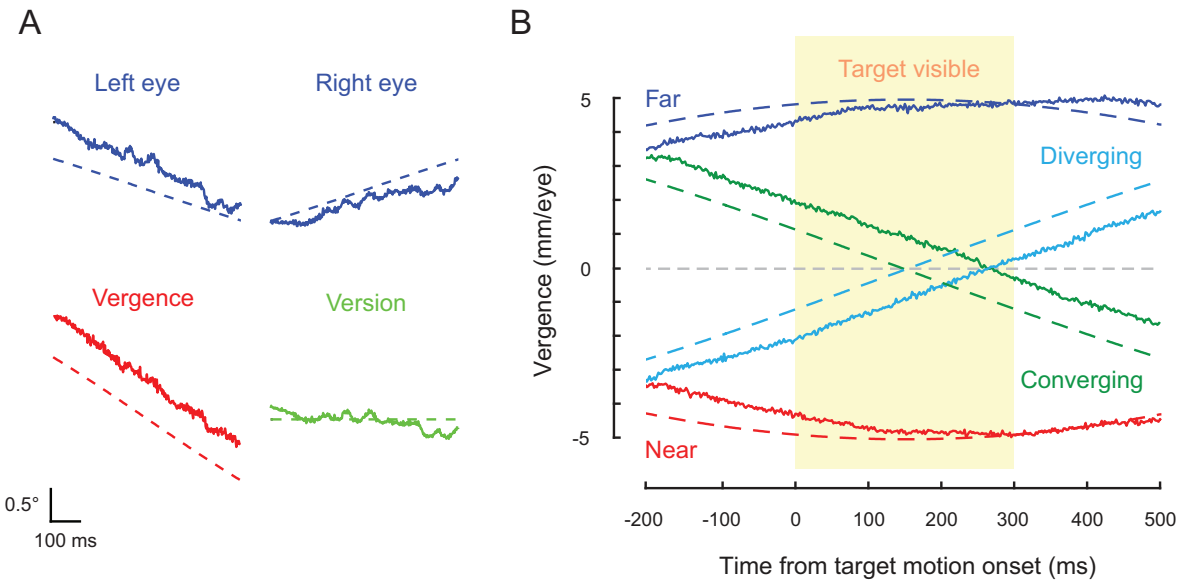

Figure 4.4: Measured eye positions from 200 ms before the target started moving until 200 ms after the target disappeared. The dashed lines indicate the background position. (A) Eye movements from a single trial in which the eyes were converging when the target was presented. Top row: left and right eye traces. Bottom row: version and vergence traces (i.e. mean of and difference between the left and right eyes traces). (B) Median vergence response for each condition (expressed in terms of half the lateral distance between where the two eyes were directed on the screen).

ity as a function of the target's motion relative to the background (Figure 4.5). Fitting these data with a cumulative Gaussian yielded the point of subjective equality (PSE; 50% point on the curve) to provide a measure of the perceived velocity of the target. We defined velocity discrimination thresholds (Δv) as the standard deviation of the fitted Gaussian and we defined increment thresholds (i.e. Weber fractions) as the ratio of threshold velocity to the mean relative velocity ($\Delta v/v$). Because individual data did not vary significantly across observers (e.g. see Figure 4.5D), we collapsed the individual data across observers. Statistical tests were performed on (uncollapsed) individual data.

We found that judgments of 3D velocity differed significantly between conditions (repeated measures ANOVA on the PSE's: $F_{3,15} = 18.5, p < .001$). The most striking feature of the data was that a target with the same retinal velocity was seen as faster during convergence than during divergence (Figure 4.4 - green vs. cyan data series). The difference between the

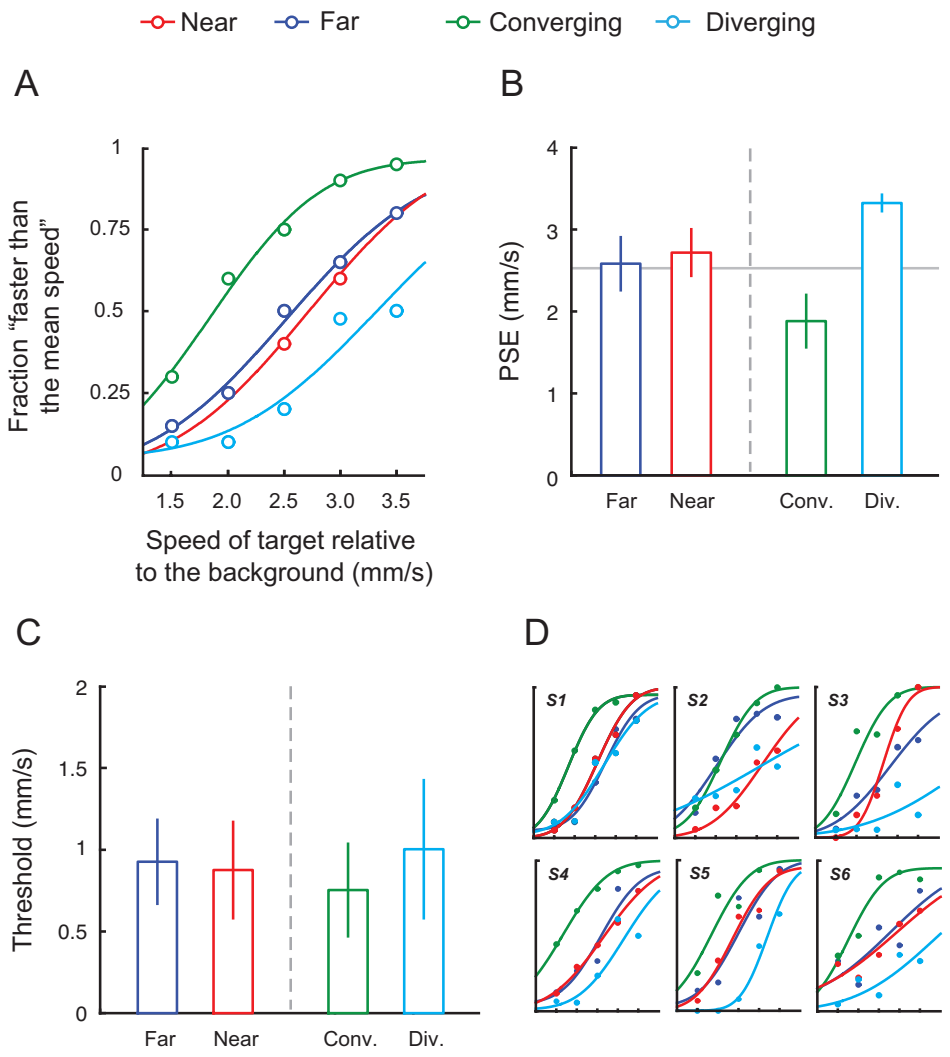

Figure 4.5: Psychophysical results. (A) Psychometric functions for the fraction of trials in which observers responded that the target approached faster than the mean approach velocity as a function of the relative velocity on the screen. Separate functions for the conditions in which the eyes were in near (red) and far (blue) vergence positions and when the eyes were converging (green) and diverging (cyan). (B) Points of subjective equality with standard errors. The solid horizontal line represents the mean velocity of the target with respect to the background. (C) Increment thresholds with standard errors. (D) Individual subjects' data.

PSEs was 1.42 mm/s of lateral velocity (consistent with a 3D velocity of about 2.2 cm/s, about 60% of the standard velocity of the target; $t_5 = 6.03, p < .01$). We found no significant shift between the psychometric functions for the nearly stationary eyes in the far and near vergence positions (the difference was 0.14 mm/s; $t_5 = .69, p = .52$).

Increment thresholds did not differ significantly between the conditions ($F_{3,15} = 2.36, p = .113$; Figure 4.4C). This result was expected, given that judgments were made with respect

to an internal standard that was the mean of the entire stimulus set. We found that the mean increment threshold for approach velocity was 0.22 (across observers and conditions). This is consistent with previously reported thresholds for the velocity of motion-in-depth (0.20; Harris & Watamaniuk, 1995) and lateral motion (0.25; McKee, 1981).

4.4 Discussion

When judging 3D motion of a target, observers tend to change vergence to track the object. To judge 3D motion correctly, the visual system should therefore take account of eye movements. Here we isolated extra-retinal cues to vergence from the retinal cues that would normally accompany vergence eye movements and tested how extra-retinal information is combined with retinal signals to judge 3D velocity. Our results show that extra-retinal cues to vergence pursuit movements systematically affected judgments of 3D velocity: an object's approach velocity is reported to be faster during convergence and slower during divergence. This must be due to the rotation of the eyes rather than to distance scaling because judgments were not sensitive to whether the eyes were at near or far vergence positions. Specifically, the transformation of a changing retinal disparity signal to 3D velocity depends on knowing the viewing distance: from binocular geometry, the same rate of change in disparity at a far distance should result in a faster perceived 3D velocity than when it is presented at a closer distance. Interestingly, in our experiment we did not find evidence for such scaling.

In the Introduction, we outlined sets of predictions for performance in these different conditions under two different models. When we compare our psychophysical results (Figure 4.5) with these models' predictions (Figure 4.3) it is clear that neither provides a good account of our results. We will consider two possible reasons for this: that the perceived velocity is determined by a combination of the two sources of information and that the perceived velocity

is scaled by ocular convergence despite the background appearing to remain at a constant distance. However, first we will briefly discuss the eye movements themselves.

4.4.1 Effects of pursuit lag

Although observers pursued the background almost perfectly, we found that the eyes lagged the background by about 100 ms (4.4B). Do we need to consider this lag when interpreting our results? Both the models we outlined in the Introduction are insensitive to lag: the relative velocity model considers the difference between the retinal velocity of the background (B) and target (T), so introducing an equal increment or decrement to both has no effect on their difference; the absolute velocity model is independent of the magnitude of the pursuit signal (P) because it only depends on target velocity (T) - lag will affect pursuit and retinal slip of the target (T') to equal and opposite extents, so their sum will not change. We can formulate modified models that consider that retinal slip is combined with a later eye movement (Rotman, Brenner, & Smeets, 2004, 2005), in which case the target will appear to move faster in the far than in the near condition, because a later eye movement will include more convergence for the far target and more divergence for the near target. We see some indication of this in Figure 4.5A, but the individual subjects' data in Figure 4.5B suggest that this is all due to one subject (S2). For the approaching and receding conditions this effect should be minimal because the velocity of the background hardly changes near the time at which the target is presented.

4.4.2 Scaling angular velocity by viewing distance

From our results it is clear that changes in the orientation of the eyes affect judgments of 3D velocity. In seeking to explain our data we have so far only considered models that use angular velocity measures. As the retinal projection of movement depends on the viewing distance, it

is reasonable to ask whether observers recovered the real-world velocity, scaling the angular retinal velocities by the viewing distance. In our displays there are two potential sources of information about the viewing distance: the vergence position of the eyes (cf. Enright, 1991; Foley, 1980; J. P. Frisby, Catherall, Porrill, & Buckley, 1997; Collett et al., 1991; Brenner & van Damme, 1998; Taroyan, Buckley, Porrill, & Frisby, 2000; Backus & Matza-Brown, 2003) and the gradient of vertical disparities in the projection of the background (Bradshaw, Glennerster, & Rogers, 1996; Backus, Banks, van Ee, & Crowell, 1999; Brenner et al., 2001). These two sources were always consistent with each other. The difference in vergence between the near and far conditions was about 1.14 deg. Changes in the vertical extent of the stimulus (i.e. the vertical separation between triangles at the top and bottom of the background in the two eyes) were 4.6 arcmin (0.2 %) at most. Note moreover that changes in vertical disparity are negligible for the target (the object the observers were judging), so vertical disparity can only contribute to scaling (or judging eye rotation on the basis of the background's retinal image deformation). Considering the data from the near and far vergence conditions allows us to assess the extent to which observers scaled angular velocity estimates to judge 3D velocity. In particular, if observers scaled the retinal velocities by the vergence distance, we would expect that the same retinal velocity is perceived as faster at the far vergence position.

Figure 4.6 shows what our data would look like if observers recovered the 3D velocity of the target by scaling the angular measurements by eye vergence. Comparing the predictions of a real-world (3D) velocity model (Figure4.6A) with our psychophysical results (Figure4.5A), it is clear that observers did not recover the real-world velocity. If we consider the individual subjects' data (Figure4.5), it is clear that only one participant (Subject S2) shows a difference between judgments in the near and far vergence positions. However, this observer's data does not match the predictions of the real-world velocity model (Figure4.6A). We conclude that

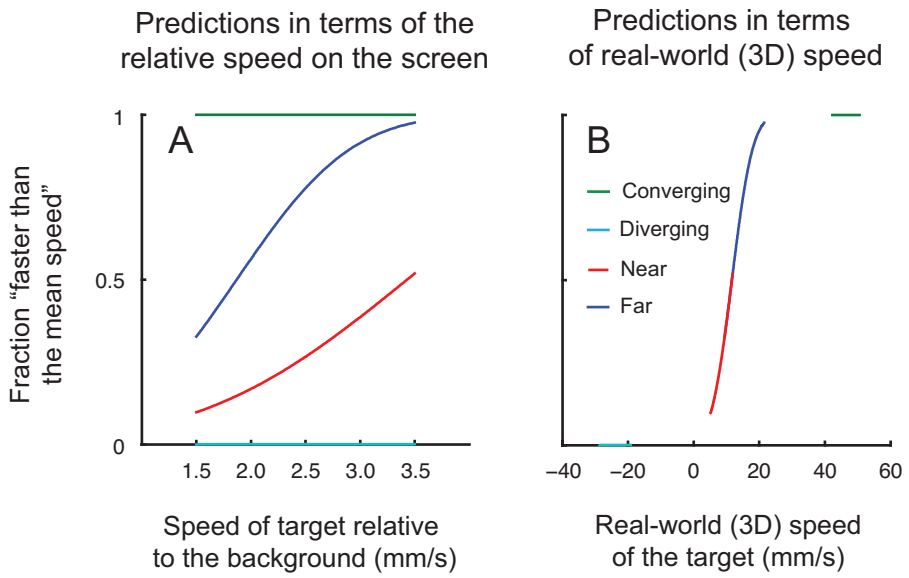

Figure 4.6: Predictions for a model that recovers the real-world velocity of the target, both in terms of the relative velocity on the screen. (A) and in terms of the real world velocity that would coincide with the simulated target positions on the screen (B). Note that this model scales the lateral motion by the viewing distance when judging velocity, in contrast to our original models.

judgments of 3D velocity are affected by extra-retinal signals about changes in eye orientation (i.e. the differences between the converging and diverging conditions), but are unaffected by extra-retinal signals relating to the baseline vergence of the eyes (i.e. the lack of significant differences between the near and far conditions). This lack of scaling conforms to previous work on the scaling of velocity in the fronto-parallel plane (McKee & Welch, 1989), suggesting that the visual system codes 3D velocity signals in angular dimensions, uncorrected for viewing distance.

4.4.3 Relative contribution of retinal and extra-retinal signals

It is clear that none of the models we have considered so far can account for our results: observers do not base their judgments on the relative velocity of the target with respect to the background or the absolute angular velocity of the target. In addition, we have shown that observers do not recover the target's real-world velocity. What estimate of velocity were our subjects using? Both of our original models exploit angular velocity measurements and both

would normally (if the background were not 'moving') provide estimates of the target's motion in depth. Observers may therefore exploit a weighted average of both signals. To test whether this could account for observers' performance, we constructed a model that calculates a weighted average of information about the retinal slip of the target relative to the background (MODEL 1; $T' - B'$) and information about the absolute velocity of the target (MODEL 2; $P + T'$). The weight term (w) determines the relative contributions of both sources of information. Specifically, we described the estimated velocity (V) as:

$$V = w(P + T') + (1 - w)(T' - B'), \tag{4.1}$$

which can be simplified to:

$$V = wP + T' - (1 - w)B'. \tag{4.2}$$

The best way to combine the two measures is for the weight term w to be chosen such that the precision of V is maximized (Cochran, 1937). This is so when w is related to the variances associated with the retinal flow of the background (σ_B^2) and of the pursuit signal (σ_P^2) by:

$$w = \frac{1/\sigma_P^2}{1/\sigma_P^2 + 1/\sigma_B^2} \tag{4.3}$$

Note that although the variance in the percept also depends on the variance associated with the retinal slip of the target (σ_T^2), the latter contribution is independent of w so it does not influence the optimal weight. Using previously published data (Experiment 1 from Welchman et al., 2009), we estimated the relative sensitivity of observers to retinal and extra-retinal signals to be a factor of 2.86. Thus, we estimate that $\sigma_P = 2.86\sigma_B$. Substituting this in equation

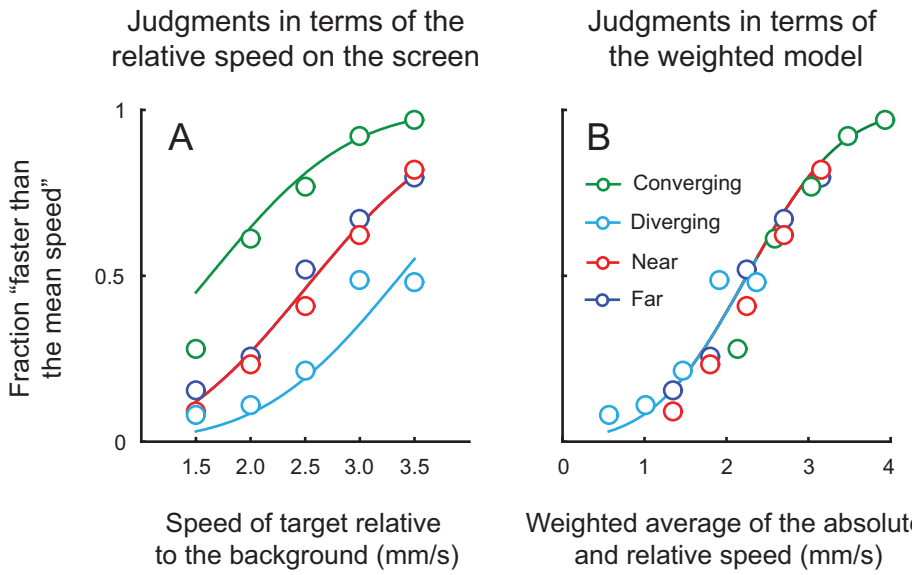

Figure 4.7: Predictions for a model based on a weighted average of the absolute velocity of the target and the velocity of the target relative to the background. The solid curves show model predictions. Data points show the actual fractions of judgments. (A) Predictions and judgments in terms of the relative velocity on the screen. (B) Predictions and judgments in terms of the weighted model. Note that the curve for far (blue) is always obscured by the curve for near (red).

4.3 gives $w = 0.11$.

This weighted model provides an excellent fit to the psychophysical results (Figure 4.7). We therefore conclude that the visual system estimates the velocity of motion in depth by taking a weighted average of changes in relative retinal disparity (Model 1) and changes in the target's angular rate of convergence as estimated from the sum of retinal slip and extra-retinal eye velocity (Model 2). This is consistent with studies that have shown that observers use extra-retinal signals when judging velocity in the fronto-parallel plane (Brenner & van den Berg, 1994; Champion & Freeman, 2010; Freeman, 2001; Freeman & Banks, 1998; Freeman, Champion, & Warren, 2010; Freeman & Fowler, 2000; Turano & Massof, 2001).The weight that we find for the extra-retinal component is lower than the extra-retinal gain terms found for the estimation of lateral motion, which are typically between 0.6 and 0.8 (Freeman & Banks, 1998; Freeman & Fowler, 2000; Turano & Massof, 2001). However, looming is likely to typically provide a substantial contribution to judgments of motion in depth, for which

there is no equivalent for lateral motion. Alternatively, the different weights may be due to the much larger angular velocity relative to each eye required for lateral motion than for a similar amount of motion in depth.

Other work also suggests a lower weight (ca. 0.4) for extra-retinal information when judging 3D motion (Howard, 2008). Moreover, the weight is likely to depend on the velocities involved. For instance, consider that under our paradigm the target appears static until it starts moving relative to the background. From Eq 2, this might appear unexpected. Specifically, the only way to estimate V=0 when there is no relative disparity and some angular motion, is to assume that the weight given to angular motion is zero during this portion of the trial. In fact, giving zero weight to the angular motion in this situation is quite reasonable as the variance associated with the estimate of velocity from relative motion is likely to be very low (i.e. the lack of relative retinal motion in the display is quite certain). For two objects moving in depth at slightly different velocities, as is the case in the critical conditions of our experiment (approaching and receding), the certainty in the estimate of their relative motion is likely to be large because the difference in velocity is small, whereas the certainty in the estimate of each of their absolute velocities is likely to be small because both objects are moving fast. In light of this, our study probably underestimates the role of extra-retinal information in judging motion in depth because relative motion was probably given more weight than would normally be the case.

4.4.4 Conclusion

In conclusion, by independently manipulating the extra-retinal signal and measuring the resulting biases in velocity perception we demonstrate that the human visual system exploits extra-retinal signals when judging the velocity of motion in depth, just as it does for judg-

ments of lateral velocity. We present a model that uses a weighted combination of changes in relative disparity and differences in angular motion to reproduce measured perceptual judgments. Finally, we show that perceived velocity is not scaled by the orientation of the eyes, indicating that under our paradigm people judge angular velocities rather than velocities in 3D space. Acknowledgments

4.4.5 Acknowledgments

This work was funded in part by the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement number 214728-2. We thank two anonymous reviewers for their comments and suggested improvements.

5

The influence of retinal size on disparity-defined distance judgments

From simple geometry, the retinal projection of an object in the environment depends on both its physical size and distance from the observer. Thus, given a sensed retinal size the brain should not know the distance of the object: the retinal measurement is compatible with infinite combinations of physical sizes and distances. Many previous studies have shown that, under monocular viewing with all other cues to depth eliminated, an object projecting a larger retinal image is perceived as closer. When other cues to depth (e.g. disparity and vergence) are available, the effect of retinal size on depth estimates should be minimal. We measured psychometric functions of disparity-defined depth of equal or differing retinal size. We find that retinal size influenced observers' judgments of distance: targets with a larger retinal image were systematically reported as being seen closer than a target with a smaller retinal size at the same disparity-define distance. In addition, we show that the amount of perceptual bias increases as (1) the ratio between the retinal sizes of the two targets increased; and (2) the absolute distance to the two targets increased. To account for our findings, we present a probabilistic model that combines binocular disparity with retinal size information, assuming that the retinal sizes of the two targets are produced by targets of the same physical size.

5.1 Introduction

Judging the distance to and between objects in our environment is a principal task of the visual system. In natural viewing, perceived depth depends on many sources of visual information (cues) that signal depth, such as binocular disparity, perspective and motion parallax. However, in going from the retinal image to depth perception, the brain is posed with a fundamental problem: the third dimension (depth) is not directly available from the two-dimensional retinal input ("inverse optics" problem). For example, binocular disparity can signal relative but not absolute distance; the interpretation of metric depth from disparity depends on the visual system combining retinal information with extra-retinal information about the orientation of the eyes.

One means of circumventing this, is for the visual system to combine multiple pieces of sensory information (depth cues) that are subject to different constraints. Most contemporary theories of cue combination are based around a modular Bayesian framework, in which each cue is first processed separately, followed by a probabilistic combination stage in which each cue's relative contribution depends on its relative reliability (e.g. Landy et al., 1995). This Bayesian approach has successfully accounted for many aspects of depth perception, including surface slant (Hillis et al., 2002, 2004; Knill & Saunders, 2003), object shape (Ernst & Banks, 2002; Johnston, 1991; Johnston et al., 1993) and three-dimensional speed (Lugtigheid, Brenner, & Welchman, In Press). However, only a few recent studies have investigated this cue combination approach in the context of distance perception (e.g. Held et al., 2010; Svarverud, Gilson, & Glennerster, 2010)

For example, Svarverud et al. (2010) showed that results from a distance-matching task in a virtual expanding room could be predicted by a weighted combination of physical cues (cues that allow reconstruction of the 3D environment, i.e. binocular disparity and motion parallax)

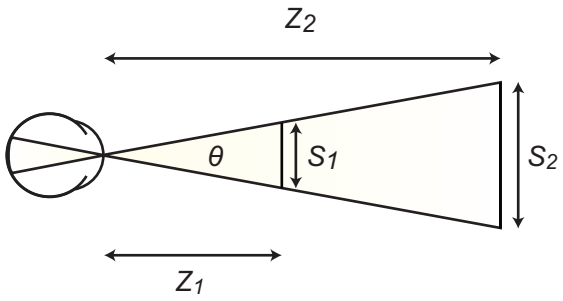

Figure 5.1: Ambiguities in distance from retinal size Retinal size cannot unambiguously signal distance: a single retinal measurement is consistent with infinite combinations of distance and size. Here, this is illustrated by two objects (1 and 2) that are located at a different physical distance (z) and have a different physical size (s), yet their projected size on the retina is equal.

and texture-based cues (cues that are invariant despite the expansion of the room) cues. Held et al. (2010) investigated how blur (image defocus) contributes to depth perception. Blur is a scale ambiguous cue to depth; i.e. on its own, blur is unable to provide a single estimate of focal or relative distance. However, Held et al. (2010) showed that when blur is probabilistically combined with other ambiguous pictorial cues (e.g. perspective), the combination of cues could provide an estimate of both focal and relative distance information.

Here we investigate how the visual system combines depth information from binocular disparity and retinal size. Like blur, binocular disparity information can only convey relative depth information, but the recovery of a metric depth estimate is constrained by the geometry of binocular vision. Specifically, correct interpretation of relative disparity depends on knowing the viewing distance and the visual system may use other cues (e.g. the convergence angle of the eyes and/or vertical disparity) to provide this viewing distance. Retinal size could provide absolute depth information when the object is familiar and its size is known. For example, artists create realistic illusions of depth in paintings by reproducing the monocular retinal image that perspective forms on our retina when observing the depicted 3D scene. However, when an object is unfamiliar the brain should not be able to recover distance information from retinal size alone: the retinal measurement is consistent with infinite combinations of physical object sizes and distances (Figure 5.1).

However, in the natural environment there is an invariant relationship between the physical size of an object and its distance, such that an increase in the object's distance results in a corresponding change in the size of its retinal projection. In a laboratory setting it can be demonstrated that - under monocular viewing and with all other cues to distance eliminated - variations in the retinal size of an unfamiliar object can lead to variations in perceived depth (for a review, see Sedgwick, 1986). Under these constrained monocular viewing conditions, retinal size has been shown to serve as a cue to distance in three ways: (i) a change in the retinal size of an object is normally interpreted as a change in its distance, rather than a change in its size (Ittleson, 1951a); (ii) absolute distance estimates vary as a function of retinal size (Over, 1963); and (iii) when two stationary objects are presented, their relative retinal size influences their perceived relative distance, the object subtending the larger visual angle appearing to be closer (Ittleson, 1951b; Hochberg & McAlister, 1955; Gogel, 1964). In general, it is known that size-matches under reduced viewing are dominated by retinal size (Gogel, 1969). When the amount of monocular distance information is increased, size judgments are more likely to be based on the actual physical size; i.e. there is size constancy (Holway & Boring, 1941). However, judgments made under monocular viewing conditions seem unrepresentative for viewing under natural conditions, in which observers have access to other cues to compensate for the ambiguity in the retinal image (e.g. disparity and vergence cues).

In a series of experiments, we investigated how monocular retinal size information interacts with binocular disparity. Geometrically, the combination of the convergence angles of the eyes with relative disparity is unambiguously related to relative distance at short viewing distances (Ono & Comerford, 1977; Foley, 1980). Thus, a priori we expect that retinal size has a negligible influence when disparity information is available. In particular, the ambiguous distance information signaled by retinal size seems unlikely to influence distance

estimates provided by relative disparity and vergence. To investigate this idea, we sequentially presented two stimuli of equal and different retinal size at the same distances. We found that differences between the retinal size of two objects systematically affected judgments of disparity-defined distance, so that an object with a larger retinal size was seen as closer than an object with a smaller retinal size at the same disparity-defined distance. To account for our results, we present a probabilistic model that combines binocular distance cues with relative size information.

5.2 General methods

5.2.1 Apparatus

Stimuli were created using OpenGL graphics libraries, implemented in the C# programming language and were rendered using anti-aliasing and geometric perspective projection from each eye. The stimuli were presented stereoscopically on a two-monitor haploscope (viewing distance 50 cm) with inter-pupillary spacing and vergence angles configured for each individual observer. The graphics card (Nvidia Quadro 4400) displayed 1600 x 1200 pixels (an individual pixel subtended 1.55 arcmin) at a refresh rate of 100 Hz.

5.2.2 Stimuli

In all experiments reported in this chapter the stimuli consisted of white outline discs against a black background. A peripheral reference volume was visible throughout the duration of the experiment to provide a reference frame for relative disparity. This reference volume formed a hollow box around the target and extended from about 30 cm to about 70 cm from the observer (Figure 5.2A). Textured cubes were placed on an invisible grid of points on the outline of the box, with each 'layer' separated by about 10 cm. The distance from the centre of the target to the background was about 12 cm (about 13.5 deg from the middle of the screen)

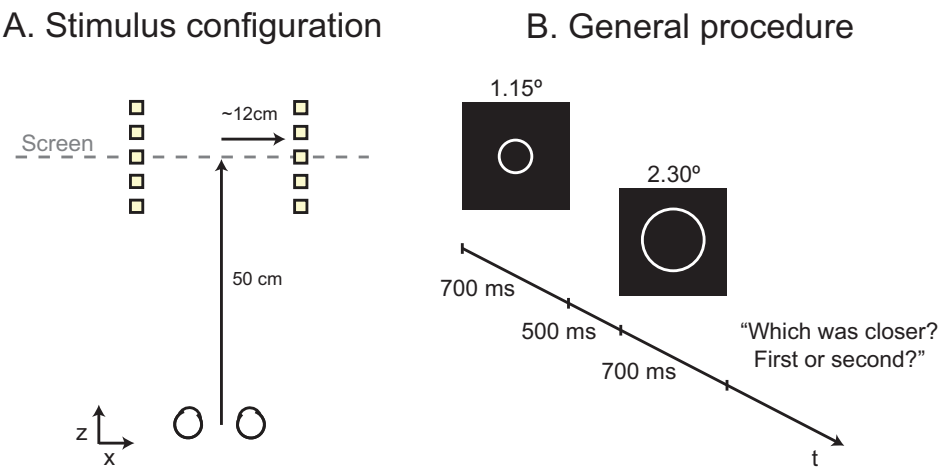

Figure 5.2: General stimulus and experimental procedures (A) Top down view of the stimulus configuration. The screen is located at 50 cm from the observer. The targets, presented in the midline between the eyes, were surrounded by a peripheral reference volume that was at a lateral distance of approximately 12 cm (about 13.5 deg). (B) Observers sequentially viewed two targets, in this example one with a small and one with a large retinal size, and were asked which one was closer: first or second?

both in the horizontal and vertical direction. We chose a fairly large separation between the targets and the reference volume to ensure that observers made comparisons between the targets, rather than sequentially comparing the relative disparity of the target with respect to the non-changing background. To avoid inappropriate matching of features in the reference volume (i.e. the wallpaper effect) some cubes were omitted to create an irregular grid.

5.2.3 Procedure

In all experiments except Experiment 4 observers matched the distance of two sequentially presented targets by indicating - by means of a key press - whether the first or second target they saw was closer (Figure 5.2B). One interval contained a standard stimulus and the other contained a comparison stimulus, the distance of which was parametrically varied around the standard using the method of constant stimuli (i.e. 2IFC). The order in which interval the standard was presented was randomised on a trial-by-trial basis. We specifically chose a sequential paradigm to provide optimal conditions to measure any effect of retinal size: if we had presented the targets simultaneously the task would essentially be reduced to a vernier

task, which would reduce our ability to measure any influence of retinal size (e.g. see Westheimer, 1979). Presentation duration was 700 ms, with a blank inter-interval pause of 500 ms. In all experiments, we obtained two PSE's per condition: one in which the small stimulus was the standard and another for which the large stimulus was the standard. After each trial, observers pressed a key to advance to the next trial.

5.2.4 Data analysis

Figure 5.3 shows the results for Experiment 1 for a single observer and provides an illustration of how we quantified our results. To obtain estimates of the PSE (50% threshold), the discrimination threshold and their respective 95% confidence intervals, we first constructed psychometric functions for the proportion of trials on which the comparison stimulus was seen as farther as a function of the disparity-defined depth separation of the standard and comparison stimuli (e.g. see Figure 5.3A-B). We then fitted a cumulative Gaussian to our data using the percentile bootstrap method implemented by the psignifit toolbox for Matlab (Wichmann & Hill, 2001; estimates were based on 1999 simulations). The standard deviation of the fitted Gaussians was divided by $\sqrt{2}$ to quantify the discrimination threshold (because we used a 2IFC paradigm).

Average bias (here described as the difference between observers' estimates of disparity-defined depth and its actual physical value) was calculated by averaging the shift in the PSE's obtained for small and large standard stimuli as follows. We reasoned that if there were a systematic shift in the PSE, we would find a directional bias, consistent with the results that have been reported under monocular viewing. Specifically, a large test would be seen at a closer distance than a small standard stimulus, and vice versa. That is, the PSE shift for a large test was expected to be negative (the large test would have to be adjusted to be farther to be

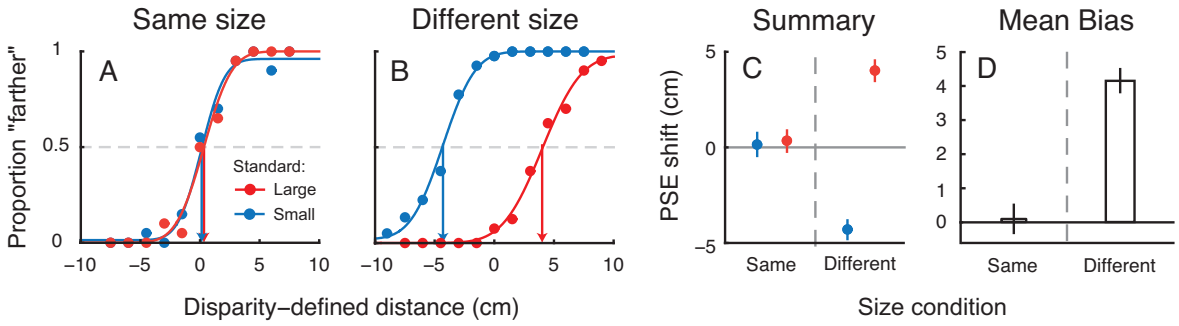

Figure 5.3: Single observer results for Experiment 1 The blue data series shows data obtained with a small reference and a large test stimulus; red data series shows data obtained with a large reference and a small test stimulus. (A) Psychometric functions for disparity-defined depth for the "same-size" condition. (B) Psychometric functions for disparity-defined depth for the "different-size" condition. (C) Summary of the results from the same- and different size conditions. (D) The mean bias, calculated by first multiplying the data collected with the small standard by -1 and then averaging these data with the data collected for with a large standard. All error bars show 95% confidence intervals.

seen as equidistant to the small standard). By the same reasoning, the PSE shift for the small test was expected to be positive (see Figure 5.3C). To quantify the average shift in the PSE - and thereby the perceptual bias - we first multiplied all 1999 bootstrap estimates for the PSE of the large standard by -1. We then averaged these bootstrap values with the 1999 bootstrap values for the PSE of the small test stimulus by adding the values in a point wise manner and dividing them by two. This resulted in a bootstrap distribution for the average estimate of the PSE with 1999 averaged values. The average bias (the median) and its associated 95% confidence intervals were then obtained from this average bootstrap distribution (see Figure 5.3D). The same procedure was used to obtain a combined estimate for the standard deviations of the fitted Gaussians.

5.2.5 Observers

The observers in each experiment were recruited from staff and students at the University of Birmingham. Each had normal or corrected-to-normal vision and could discriminate at least 1 arcmin of relative disparity in a briefly presented random-dot stereogram. Appendix B of this thesis provides the details of observer screening.

5.3 Experiment 1

5.3.1 Methods

In the first experiment, we establish how retinal size and relative disparity interact. We measured psychometric functions for disparity-defined distance in five interleaved conditions. In the first condition observers sequentially viewed pairs of small (7 arcmin) Gaussian dots ("no-size" condition); In the second condition observers viewed pairs of outline circles of equal retinal size ("same-size" condition: large-large and small-small); In the third condition observers saw pairs of outline discs of unequal retinal size ("different-size" condition: large-small and small-large; i.e. one condition in which the large target was the standard and one in which the small target was the standard). Large targets had a retinal size of 2.3 deg, small targets had a retinal size of 1.15 deg (i.e. their retinal size ratio was 2). The disparity-defined distance of the standard stimulus was kept constant at 50 cm; the distance of the test stimulus was systematically varied around the standard's distance by various amounts, with the range of measurements tailored to each subject. This ensured that we could measure a full psychometric function for all conditions despite there being large inter-observer differences in perceptual bias and sensitivity. Nine naïve observers participated in this experiment (mean age = 22.4±2.6 years).

5.3.2 Results

Our results show that retinal size significantly affected judgments (repeated measures ANOVA with 'size' as the within-subjects factor: $F_{2,16} = 44.28, p < .001$; Figure 5.4). In particular, we found a significant shift in the PSE (bias) when observers saw two object of different retinal size (single sample t-test with respect to zero: $t_8 = 6.73; p < .01$; Figure 5.4A), such that a large target was on average seen as closer than a small target and a small target was seen as

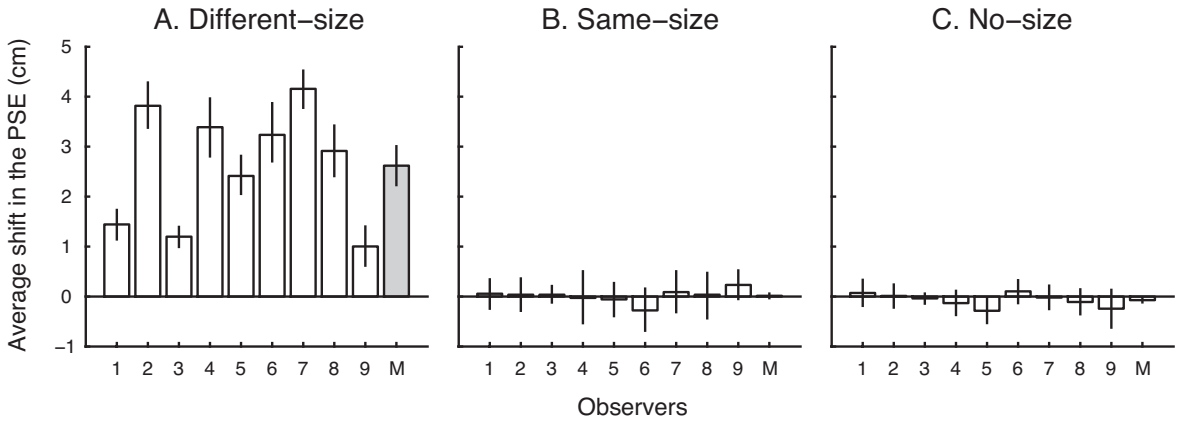

Figure 5.4: Results of Experiment 1 for nine observers. Panels show the average shift in the PSE for (A) the different-size condition; (B) the same-size condition; and (C) the no-size condition. Error bars show 95% confidence intervals. The grey bars (labeled 'M') show the mean bias across observers with the SEM.

farther than a large target. On average, the shift in the PSE was 2.6 cm (23.61 arcmin at 50 cm). Large differences were found between observers' perceptual bias ($F_{1,8} = 43.71, p < .001$; Figure 5.4A) ranging from about 1 cm (9 arcmin) to about 4 cm (36 arcmin). In contrast, we found no significant shift in the PSE (bias) when observers saw two stimuli of equal retinal size ($t_8 = .31; p = .78$; Figure 5.4B) or when they viewed two small Gaussian dots ($t_8 = -1.62, p = .15$; Figure 5.4C).

Retinal size also significantly affected discrimination thresholds ($F_{2,16} = 9.92, p < .01$; Figure 5.5). Specifically, we found that threshold increased significantly ($p < 0.01$) when we moved from a simple point (mean threshold = 0.84 cm or 5.33 arcmin) to two same-sized discs (1.12 cm or 9.96 arcmin); presenting two different-sized discs slightly elevated thresholds (mean threshold = 1.38 cm or 12.16 arcmin; $p < 0.05$). The thresholds reported here are comparable to previous work on sequential stereopsis (Brenner & van Damme, 1998; Enright, 1991; Taroyan et al., 2000). The elevated thresholds of the same-size condition relative to the small point is expected, as there is a small cue conflict between the depth specified by retinal size and relative disparity: the retinal size of two targets in the same-size condition is constant

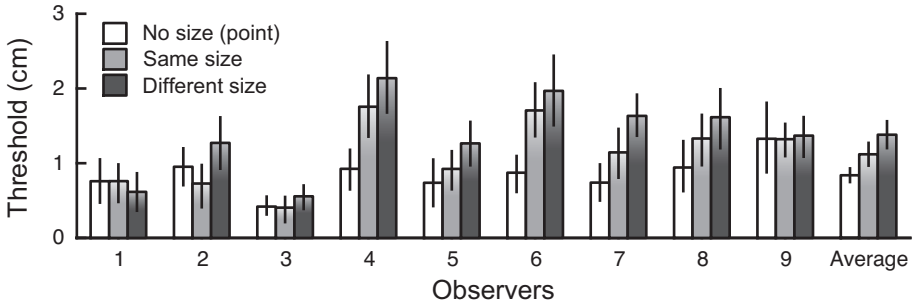

Figure 5.5: Thresholds in Experiment for nine observers in the no-size condition (white bars); the same-size condition (light grey bars); and the different size condition (dark grey bars). All error bars show 95% confidence intervals, except those for the average whose values indicate the SEM across observers.

and therefore does not signal a change in distance, whereas the relative disparity does. This may make it more challenging to discriminate the depth difference. By the same reasoning, the difference between thresholds for the same- and different-size conditions is presumably due to the larger conflict between retinal size and relative disparity; in the different-size condition, retinal size specifies a much larger depth than that specified by relative disparity and this larger cue-conflict situation may cause higher thresholds.

5.4 Experiment 2

5.4.1 Rationale

Having established that retinal size affects judgments of disparity-defined distance, we next ask how observers' judgments vary as a function of retinal size ratio. In particular, it is unclear whether this bias is due to the targets' absolute retinal size difference or their retinal size ratio. In this experiment we investigate this question using two conditions (tested in separate sessions).

5.4.2 Methods

The procedure was the same as in Experiment 1. However, in the first condition we measured psychometric functions of disparity-defined distance of small and large targets by keeping the

retinal size ratio between the two targets constant at 2. That is, we varied the absolute retinal size of the small target in four equal steps from 0.4 to 1.6 deg; the large target was twice as large as the small target (see 5.6, left bottom panel underneath the axis). In the second condition, we always presented one small target (1.15 deg) and varied the retinal size of the other in five equally spaced steps from 1.38 to 2.3 deg, such that the retinal size ratio between the two targets varied between 1.2 to 2 (see 5.6D, right bottom panel). As in the previous experiment, the distance of the standard stimulus was constant at 50 cm, the distance of the test stimulus was varied around the standard so that a full psychometric function was sampled. Six naive observers took part (mean age = 23.9±2.8 years).

5.4.3 Results

When we kept the retinal size ratio constant, we found that bias did not differ significantly between different absolute retinal sizes (Repeated measures ANOVA on the PSE's; $F_{3,18} = 1.96, p = .16$; Figure 5.6B). However, when we varied the retinal size ratio, we found that bias varied between different retinal size ratio's ($F_{4,24} = 4.59, p < .01$; Figure 5.6A). Interestingly, even though we used the same observers in both conditions, we found a difference between the magnitudes of bias. Specifically, when we kept the retinal size ratio constant at a value of two, we found that average bias was about 2.88 cm. Yet, when we varied the retinal size ratio, the bias associated with a retinal size ratio of two was about 1.71 cm. One possible explanation for this difference is that when we varied the size ratio, the standard was always the same. This may encourage the assumption that it is the same object, resulting in a larger bias. In the other case, the changes in the size of the standard are large, so this may have encouraged observers to believe that relative size is less indicative of distance.

When we varied the retinal size ratio, the change in ratio affected thresholds ($F_{3,18} =$

$3.18, p < .05$; Figure 5.6C) in a near-linear fashion, with a minimum of about 1 cm and a maximum of about 1.8 cm. When we kept the retinal size ratio constant, thresholds were relatively stable (with an average threshold of about 1 cm): the change in the absolute sizes of the targets did not affect threshold ($F_{3,18} = 3.39, p = .11$; Figure 5.6D). Finally, for all the analyses described above, we found that for both the PSE shift and for its associated threshold differed significantly between observers (all $F > 4$, $p < 0.01$; Repeated Measures ANOVA).

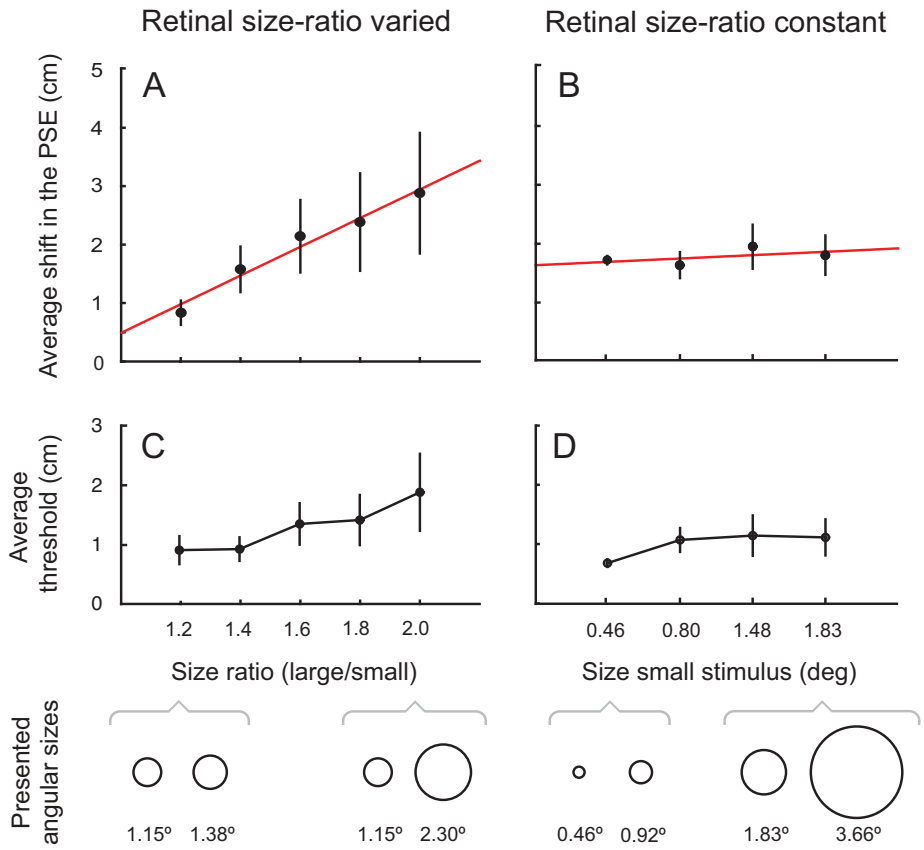

Figure 5.6: Perceptual bias as a function of retinal size ratio (panels A and C) and absolute retinal sizes (panels B and D). Panels (A) and (C) show average shifts in the PSE and average thresholds, respectively, for when we varied the ratio between the two targets' retinal sizes. Panels (B) and (D) show the average shifts in the PSE and average thresholds, respectively, for when we varied the absolute sizes of the targets, but kept their retinal size ratio constant. The red curves in (A) and (B) show regression curves. Error bars show the SEM. The stimulus manipulation (retinal size ratio varied or constant) is illustrated underneath the axes.

5.5 Experiment 3

5.5.1 Rationale

In the introduction we stated that the visual system integrates information from several cues to depth, where the influence of each individual cue depends on each cue's estimated relative reliability (e.g. see Landy et al., 1995; Ernst & Banks, 2002). This experiment tests the idea that the visual system combines information from disparity and retinal size according to their relative reliabilities. Specifically, the quality of depth information varies with distance (Cutting & Vishton, 1995; Tresilian & Mon-Williams, 2000): e.g. disparity information is most effective at near distances, but decreases as the absolute (viewing) distance increases (Howard & Rogers, 2002) whereas monocular cues (e.g. relative size) are effective at both near and far distances. Increasing the absolute distance to a point will reduce the reliability of the relative disparity signal and, as a result, its weight relative to the weight assigned to the retinal size cue. As a consequence, we would expect a higher extent of perceptual bias as we increase the absolute distance.

5.5.2 Methods

The procedure and stimuli were the same as in Experiment 1, but here we measured psychometric functions for disparity-defined depth at four different absolute distances, namely at 40, 50, 70 and 100 cm from the observer. The reference volume moved along with the change in absolute distance, under perspective projection. As a result, the retinal size of both the targets and the elements in the reference volume changed accordingly to the absolute distance. That is, at an absolute distance of 50 cm the retinal sizes of the small and large targets were the same as in Experiment 1: 1.15 and 2.3 deg, respectively. At an absolute distance of 100 cm the retinal sizes of the small and large targets were 0.57 and 1.15 deg. Seven naive observers

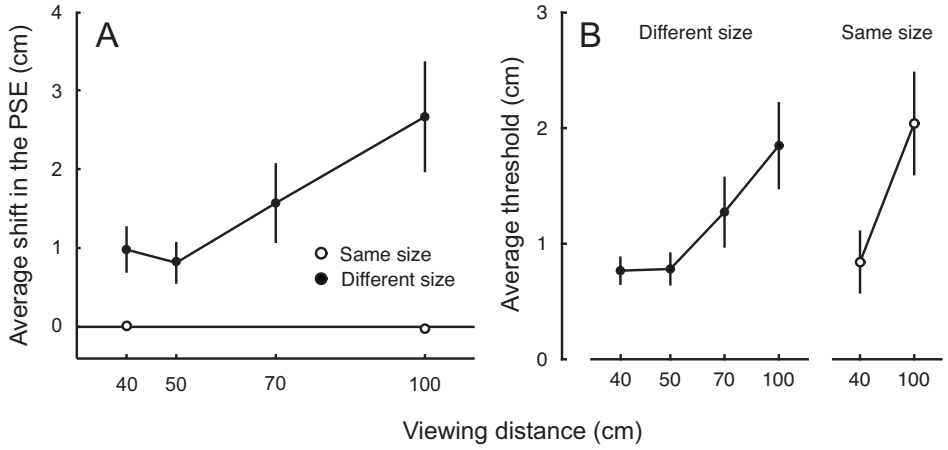

Figure 5.7: Perceptual bias as a function of changes in the absolute viewing distance. (A) Average shift in the PSE as a function of the viewing distance (40, 50, 70 and 100 cm). (B) Average threshold as a function of viewing distance. Open symbols show data from trials in which the retinal size ratio between targets was one; closed symbols show data from trials in which the retinal size ratio was two. Error bars show the SEM.

participated in this experiment (mean age = 20.2±1.3 years).

5.5.3 Results

We found that the absolute distance affected the amount of perceptual bias ($F_{3,18} = 8.73, p < .001$; Figure 5.7A, filled data series). Conversely, we found no bias when we presented same-sized objects at different viewing distances ($t_6 = .11, p = .21$; Figure 5.7A, open data series). Thresholds differed significantly between absolute distances ($F_{3,18} = 9.76, p < .001$; Figure 5.7B, filled data series) and observers ($F_{1,6} = 13.6, p < .01$). Thresholds increase as we increase the absolute distance. This increase in thresholds was similar to the increase of threshold for targets of equal retinal size ($t_6 = 4.31, p < .01$; Figure 5.7B, open data series).

Summarising, we found that bias increases as a function increasing viewing distance, providing some indication that each cue's relative contribution may be governed by their reliabilities. Increasing viewing distance seems to decrease the relative reliability of the disparity information, such that the retinal size information can exert a larger influence on judgments.

5.6 Experiment 4

5.6.1 Rationale

Binocular vision requires that the convergence angle is adjusted for proper fusion of the two retinal images, so that the point of fixation is projected onto the fovea (i.e. the high resolution central area of the retina) in each eye. In all previous experiments, we assumed that our observers' eyes were correctly converged on the target location. However, it is known that binocular fixation is typically not accurate (e.g. Ames & Gliddon, 1928; Ogle, 1950). When the difference between the vergence demand and the vergence response is large, it is known to affect perceived image size and distance (Wheatstone, 1852; Heinemann, Tulving, & Nachmias, 1959; Komoda & Ono, 1974). Specifically, the perceived size and distance of an object are reduced when the eyes are more converged than is required, an effect generally known as convergence micropsia. If convergence angles for objects of small and large retinal size differ, this could potentially account for our results. In particular, if observers' eyes are more converged for large objects than for small objects when they are at the same simulated distance, extra-retinal information to the orientation of the eyes signals a closer distance of the large object and vice versa.

5.6.2 Methods

To investigate whether convergence angles for large and small objects differed, we used the same procedure as Experiment 1 but presented a single large or small target on each trial while we measured observers' eye movements. We measured the vergence response by varying the target's vergence demand. Although unlikely to affect judgments of distance, we also measured the change in pupil size as observers viewed large and small targets. We varied the vergence demand of the targets from -1.4 to 1.4 deg (or -0.7 to 0.7 deg in each eye, respec-

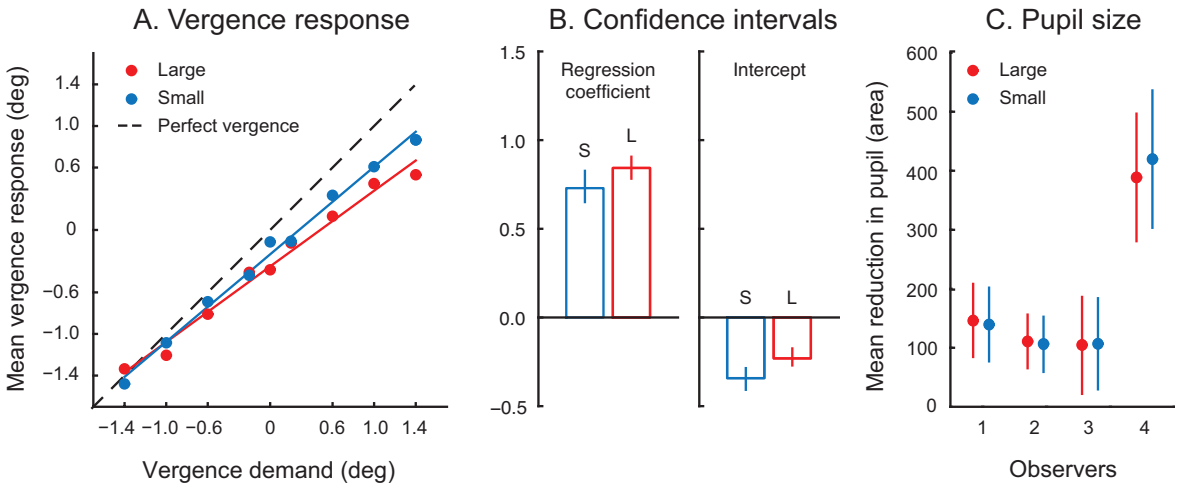

Figure 5.8: Oculomotor responses to large (red data series) and small (blue data series) stimuli. (A) the average vergence response as a function of the vergence demand (how much vergence would be required to converge on the target). The data points show the measured vergence response at each level of vergence demand we presented - the solid line represents the respective regression line. (B) Comparison of the regression coefficients and intercepts from the regression analysis in (A). Error bars show 95% confidence intervals obtained from bootstraps. (C) changes in pupil size as a function of target size. Note that the units here are unnormalised units obtained from the Eyelink system and do not correspond quantitatively to metric units.

tively), where the mean of zero was consistent with a vergence distance of 50 cm. A vergence demand of -1.4 deg was consistent with an absolute distance of about 42 cm; a vergence demand of 1.4 deg was consistent with an absolute distance of about 66.5 cm. Large and small targets were randomly interleaved. Targets were shown for 700 ms, after which there was a 500 ms pause after which the next trial started automatically. Observers were instructed to fixate the targets, but no psychophysical judgment of depth was required. Eye movements were measured using an Eyelink II (SR Research Ltd.) at a sampling rate of 500 Hz.

5.6.3 Results

Figure 5.8A shows the vergence response as a function of the vergence demand. The black diagonal dashed line shows unity, i.e. the line on which vergence responses should lie if the vergence response exactly matched the vergence demand. The most striking feature in the data is that the vergence response did not match the vergence demand. Specifically, both the vergence response for the small targets (blue data series) and the large targets (red data series)

are below the actual vergence demand and this "under-vergence" increases with higher vergence demand. In addition, there is a marginally significant difference between the vergence response to large and small targets ($F_{1,3} = 14.87, p = .048$; also see the bootstrapped 95% confidence intervals in Figure 5.8B) but this difference would not be enough to explain the perceptual bias that we find. Specifically, the maximum difference in the vergence response to small and large targets was about 0.3 deg (0.15 deg in each eye). This is consistent with a difference in the fixation distance of about 0.9 cm (at a viewing distance of 50 cm). Moreover, our results show that this difference is only relevant for higher vergence demands that lie beyond the range in which we conducted our measurements of Experiments 1 and 2. The change in pupil size elicited by the large target was not significantly different from the change in pupil size for small targets ($t_3 = .62, p = .58$; Figure 5.8C). We therefore conclude that we can rule out oculomotor factors as an explanation for our results.

5.7 Discussion

Our results show that retinal size affected judgments of disparity-defined depth: when we presented two targets of equal retinal size observers accurately judged the depth between them. However, when the retinal size differed between the two targets, judgments were biased so that a target subtending a larger retinal angle was seen as closer than a smaller target at the same disparity-defined distance. We found that the extent of this perceptual bias depended on the ratio between the retinal sizes of the objects: when we increased the ratio between the two retinal sizes, bias increased proportionally. Conversely, when we increased the absolute sizes of the two targets but kept their retinal size ratio constant, bias remained constant. In addition, we found that increasing viewing distance resulted in an increase in perceptual bias. Taken together, our results suggest that the visual system combines depth information from

disparity and retinal size. This combination process results in perceptual bias even though there is - strictly speaking - no cue conflict. Cue combination seems a curious strategy for the visual system to take, considering the ambiguous nature of the retinal size information and the perceptual bias it seems to cause. As purely disparity-based depth judgments (i.e. the "no size" condition) were unbiased, it may seem more sensible to discard (veto) the retinal size information in favor of a more accurate depth estimate from disparity and vergence (cf. Landy et al., 1995). Surprisingly, we find that this is not the case: it is clear that retinal size is taken into account when judging the depth in our experiment.

We next present a probabilistic model to account for these results. Our purpose is to provide a qualitative (descriptive) explanation of our results, considering the constraints and properties of each source of information. We first explore how the visual system may estimate the disparity-defined depth from combining relative disparity information with an estimate of the absolute distance derived from the convergence angles of the eyes. In this model, retinal size information is not included as a source of information. Rather, this primary model serves to provide a baseline measurement of depth that relates to the results of the "no-size" condition of Experiment 1, in which the targets were Gaussian dots that do not convey a relative size signal. Expanding this baseline model, we then investigate (1) how the inclusion of retinal size information results in perceptual bias and (2) how this bias varies as a function of the retinal size ratio between two targets and as a function of the absolute distance.

5.7.1 How might the visual system recover disparity-defined depth?

To explain how the visual system combines depth information from retinal size and binocular disparity, we first consider how the visual system may estimate depth from the sources of

information that were available in Experiment 1's "no-size" condition: i.e. disparity and vergence, without retinal size cues. In principle, these two geometric sources of information are sufficient to estimate depth: disparity provides (scale-ambiguous) relative depth information and vergence provides the necessary absolute distance information that can be used to scale relative disparity (e.g. see Ono & Comerford, 1977; Foley, 1980; Tresilian & Mon-Williams, 2000; Mon-Williams, Tresilian, & Roberts, 2000). However, the qualitative difference in the type of information conveyed by each cue poses a problem for traditional cue combination approaches in which a weighted average is calculated for the two cues, and the reliability of each cue controls their relative influence on the combined depth estimate. Specifically, the combination process requires each cue to provide an estimate of absolute distance (cf. Landy et al., 1995).

Thus, to combine vergence and disparity it is necessary to first express the distance information from the relative disparity and vergence information in the same units. Using the binocular geometry (see Figure 5.9A), we can specify the relative disparity between two objects (δ) at different distances (z_0 and z_1) as follows:

$$\delta = \frac{i}{z_0} - \frac{i}{z_1}, \tag{5.1}$$

where i is the inter-pupillary separation. Extracting i and multiplying by z_0/z_0 yields:

$$\delta = \left(\frac{i}{z_0}\right)\left(1 - \frac{z_0}{z_1}\right). \tag{5.2}$$

For convenience, we substitute z_1/z_0 by r, yielding an expression of disparity as a function of

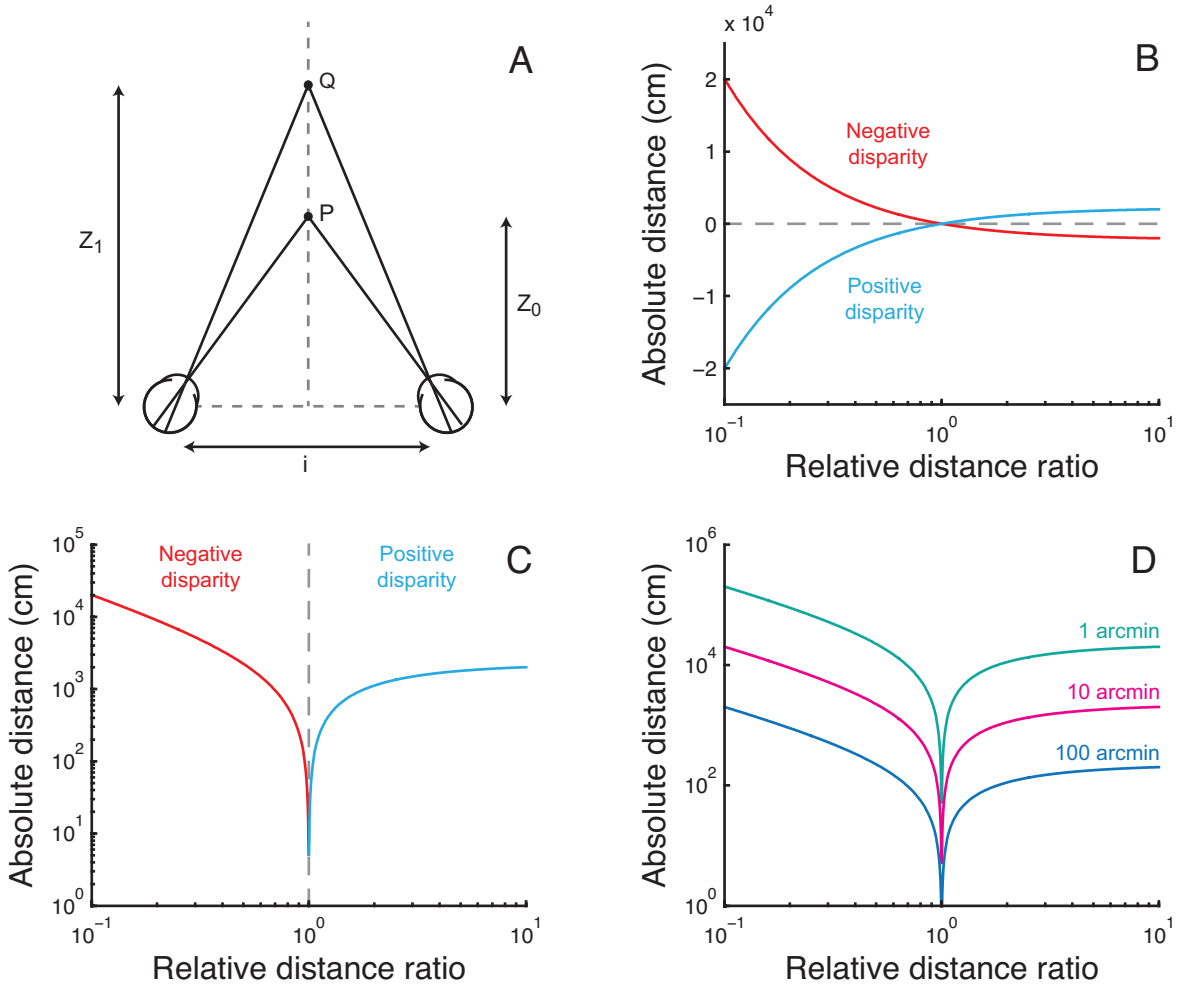

Figure 5.9: Definition of disparity as a function of relative and absolute distance. (A) Binocular viewing geometry: Point P and point Q are located at different distances (Z_0 and Z_1, respectively), with i representing the observer's inter-ocular separation. (B) The relationship between relative distance ratio and absolute distance (as shown in Eqn 5.4 for a negative value of disparity (-10 arcmin; red curve) and a positive value of disparity (+10 arcmin; blue curve). (C) The same as panel (B) but with the absolute distance plotted on a log scale: this ensures that negative values for the absolute distance are discarded. This results in a single signed characteristic 'winged' function: the red part shows the curve for negative disparity, the blue part shows the function for positive disparity. Note that the disparity sign is linked to the relative distance ratio: e.g. negative disparity results in a relative size ratio smaller than 1 and v.v. (D) Different values of disparity result in different ranges of absolute distances.

absolute and relative distance:

$$\delta = f(z_0, r) = \left(\frac{i}{z_0}\right)\left(1 - \frac{1}{r}\right), \tag{5.3}$$

where r now denotes the relative distance ratio between the two objects and z_0 is the absolute

distance. Put differently, disparity is a function of the objects' absolute distance and their relative distance ratio. However, we could in principle recover an estimate for absolute and relative distance by taking the inverse of this function and thus writing absolute and relative distance as a function of disparity:

$$z_0, r = f^{-1}(\delta) = \left(\frac{i}{\delta}\right)\left(1 - \frac{1}{r}\right) \tag{5.4}$$

Figure 5.9B plots absolute distance as a function of relative distance for two values of disparity, one positive (10 arcmin; blue curve) and one negative (-10 arcmin; red curve). Their functions differ, but are a mirror image of each other with respect to an absolute distance of zero (the position of the observer). Here, absolute distances smaller than zero would result in objects that are behind us, thus we take the absolute for the absolute distance and we plot the absolute distance on a log scale. This produces a 'wing-shaped' function as shown in Figure 5.9C. Functions for large and small disparities differ (Figure 5.9D): larger disparities (e.g. 100 arcmin; blue curve) are consistent with a range of closer absolute distances, whereas smaller disparities (e.g. 1 arcmin; green curve) are consistent with a range of farther absolute distances. However, one cannot estimate absolute or relative distance from a given disparity measurement: for each value of disparity infinite combinations of absolute and relative distances are possible.

Figure 5.10A shows the part (marked in red) of the function that we define as the 'region of interest': this region spans an absolute distance of 10 cm to 1000 cm and thus fully includes the distance range in which we conducted our experiments. We can now generate a probability density function (PDF) for depth from a measured disparity, as a function of the absolute distance (z_0) and relative distance ratio (r). To implement this PDF, we make a number of

simplifying assumptions. First, we assume that the visual system's estimate of the relative distance ratio is affected by random noise. Second, we assume that this noise is Gaussian in the log-domain. This assumption is necessary due to the lack of a closed-form solution for the respective distributions generated by disparity and vergence. However, converting a normally distributed angular measurement to a measurement of distance the resulting distribution of likely distances is skewed, in a similar manner to a log-normal distribution. We therefore propose that this is a reasonable approximation of the distributions involved in this model.

The exact value of the standard deviations of the relative distance ratio from disparity were not motivated by the experimental thresholds. Rather, we assumed that these standard deviations were approximately equal and we chose one value that provided a reasonable model fits to all conditions of Experiment 1. We finally assumed that the standard deviation of the relative distance ratio from disparity changes as a function of the absolute distance: at a farther absolute distance the estimate has a higher standard deviation, and at lower absolute distances it has a lower standard deviation. Based on the results in experiment 3 (in which we measured psychometric functions for disparity-defined depth at difference absolute distances) this seems a reasonable assumption: we found that observers' thresholds for the "same-size" condition increased as we increased the absolute distance to the two targets. The increase in the standard deviation from an absolute distance of 50 cm to 100 cm was set to 15% of the value at 50 cm.

Figure 5.10B shows the resulting probability density function for relative disparity within the region of interest marked in Figure 5.10A. The saturation indicates the probability: darker shades depict higher probability values. We can also implement a PDF for the vergence distance: this PDF is essentially linear on the absolute distance axis because vergence can only specify a single absolute distance but not relative distance (Figure 5.10C). We can now com-

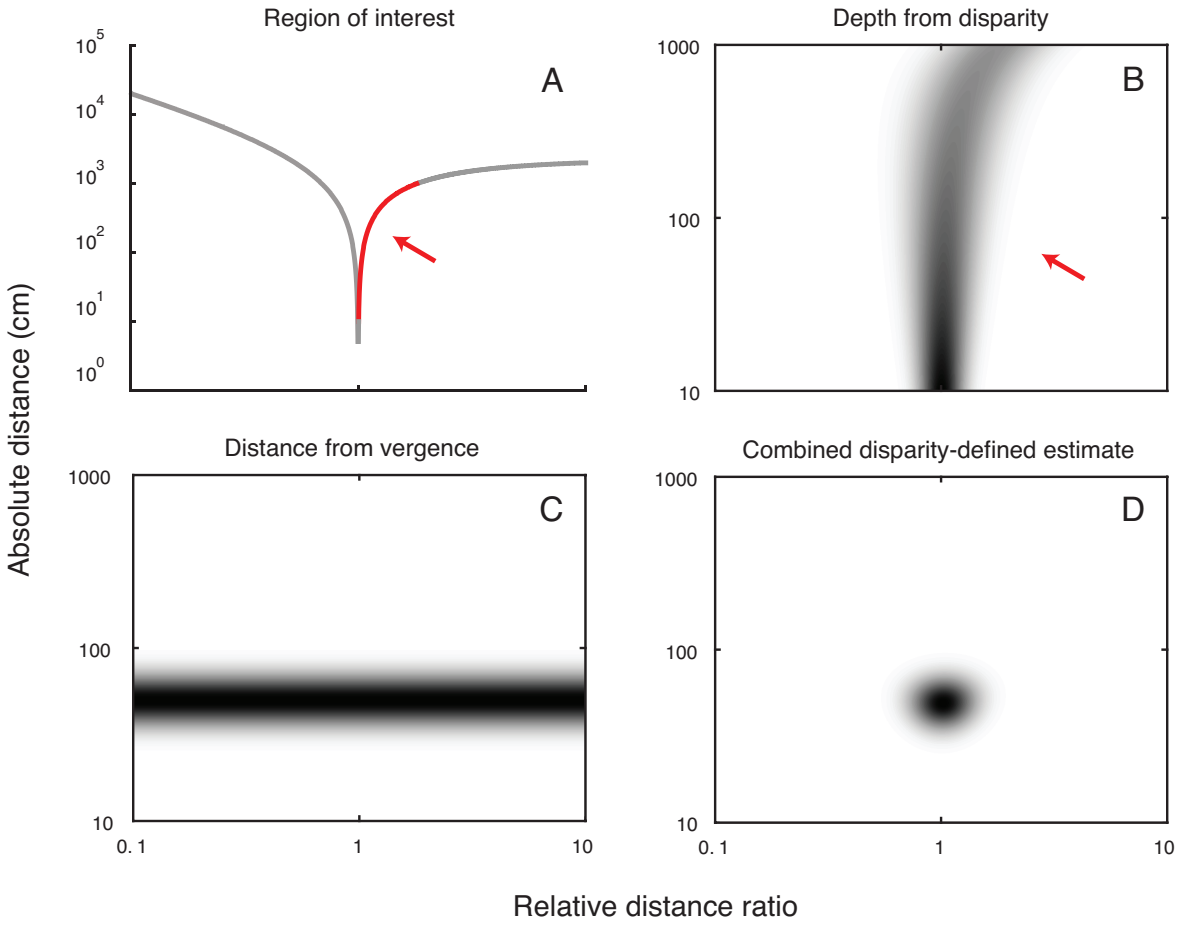

Figure 5.10: Probabilistic model of the combination of depth from disparity and the distance from vergence. (A) The red curve specifies a region of interest on the disparity curve for a single value of disparity (here +10 arcmin). (B) the probability distribution $P(z_0, r|\delta)$, where δ is the measured disparity. (C) the probability distribution $P(z_0, r|\gamma)$, where γ is the measured vergence distance. Vergence specifies the absolute but not the relative distance. (D) the product of the distributions in (B) and (C). From this posterior distribution $P(z_0, r|\delta, \gamma)$ the absolute and relative distances can be estimated.

bine the vergence distance and the relative disparity signal into disparity-defined depth by multiplying the distance-from-vergence and depth-from-disparity PDF's using Bayes' Rule:

$$P(z_0, r|\delta, \gamma) = \frac{P(\delta|z_0, r)P(\gamma|z_0, r)P(z_0, r)}{P(\delta, \gamma)}, \tag{5.5}$$

where δ and γ represent the measured relative disparity and vergence distance, respectively. $P(z_0, r|\delta, \gamma)$ is the posterior distribution for the combined estimate from disparity and vergence. Using the product rule, a technique described in Burge, Fowlkes, and Banks (2010), we

combine $P(\gamma|z_0, r)$ and $P(z_0, r)$ into the joint probability distribution $P(\gamma, z_0, r)$ as follows:

$$P(z_0, r|\delta, \gamma) = \frac{P(\delta|z_0, r)P(\gamma, z_0, r)}{P(\delta, \gamma)}, \tag{5.6}$$

Again, using the product rule, we rewrite the equation such that we obtain posterior distributions for each cue:

$$P(z_0, r|\delta, \gamma) = \frac{1}{P(\delta, \gamma)/P(\delta)} P(\delta|z_0, r)P(z_0, r|\gamma), \tag{5.7}$$

where $1/(P(\delta, \gamma)/P(\delta))$ is a normalizing constant. Note that the prior, $P(z_0, r)$, has been absorbed by $P(z_0, r|\gamma)$. We have now expressed the likelihood of a value of disparity given a value of absolute and relative distance, i.e. $P(\delta|z_0, r)$. However, in order to multiply this likelihood distribution with the likelihood distribution for the vergence distance, we would require this distribution to be expressed in the same manner; i.e. as the likelihood of the absolute and relative distances, given a value of disparity. We can accomplish this by substituting $P(f^{-1}(\delta)|z_0, r)$ for $P(\delta|z_0, r)$, yielding:

$$P(z_0, r|\delta, \gamma) = \frac{1}{P(\delta, \gamma)/P(\delta)} P(z_0, r|\delta)P(z_0, r|\gamma). \tag{5.8}$$

We can now multiply the depth-from-disparity PDF, $P(z_0, r|\delta)$, and the distance-from-disparity PDF, $P(z_0, r|\gamma)$, to obtain the estimate of disparity-defined depth (because we now have two distributions that are specified as a function of relative and absolute distance: $f^{-1}(\delta)$ maps disparity onto relative and absolute distance). Figure 5.10D shows the result of this combination process. The final disparity-defined depth estimate (Δ) is recovered from the maximum likelihood estimate (MLE); i.e. the absolute and relative distance ratio's that are most likely

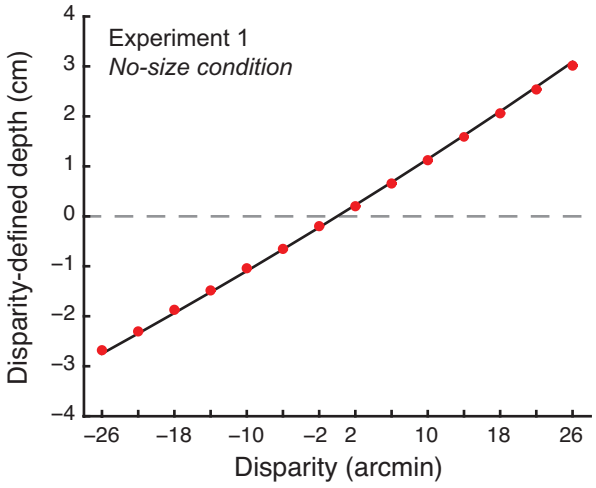

Figure 5.11: Model predictions for the disparity-defined depth over a range of disparities (±26 arcmin) for the "No-size" condition in Experiment 1. The black line represents the actual depth between the two targets as specified by disparity at a viewing distance of 50 cm. The red data points show the results of model simulations.

given the available depth information. In Figure 5.10D, this is consistent with the darkest point in the PDF. Using these most likely values of absolute and relative distance ratio, we can then recover metric depth between two targets (Δ) as follows:

$$\Delta = (r - 1)z_0 \tag{5.9}$$

Figure 5.11 (red data series) shows the disparity-defined depth estimates that are predicted by this model for a range of disparity (±26 arcmin) measurements at a viewing distance of 50 cm; the solid line shows the true depth between the two targets. It is clear that, consistent with the results from Experiment 1 ("no size" condition), the model accurately predicts the depth signaled from the combination of vergence and relative disparity: the model predicts that judgments are unbiased over a range of disparity measurements. Note that, due to Eqn 4, we cannot calculate disparity-defined depth for a value of zero (where the objects are at the same depth): we cannot divide a scalar (i) by zero. We assume that for a disparity value of zero, the disparity-defined depth is zero.

5.7.2 How does retinal size information affect disparity-defined depth?

We next explore how differences in retinal size may affect judgments of this (baseline) disparity-defined depth. As outlined in the Introduction, the brain should not know the distance of a single unfamiliar object from its retinal size: the retinal measurement (α) can be consistent with infinite combinations of physical sizes (s) and distances (z):

$$tan\alpha = \frac{s}{z} \tag{5.10}$$

The same premise is true for the comparison between two objects (0 and 1):

$$\frac{tan\alpha_0}{tan\alpha_1} = \frac{s_0}{s_1}\frac{z_1}{z_0} \tag{5.11}$$

A single ratio of retinal sizes cannot uniquely specify the distance ratio, as the retinal size ratio can be consistent with infinite combinations of physical size ratio's and distance ratio's. However, under some circumstances the retinal size ratio of two unfamiliar objects may provide some depth information. Specifically, when it is assumed that they are of equal physical size (i.e. $s_0 = s_1$) their retinal size ratio is inversely proportional to the ratio of their physical distances (cf. Gogel, 1969):

$$\frac{tan\alpha_0}{tan\alpha_1} = \frac{z_1}{z_0} \tag{5.12}$$

Again, it is convenient to substitute r for z_1/z_0, yielding:

$$\frac{tan\alpha_0}{tan\alpha_1} = r \tag{5.13}$$

In other words, under this assumption with a retinal size ratio of two (i.e. object 0 has a retinal size that is twice as large than object 1) the monocular information specifies that the distance to the object at z_1 must be twice as far than the object at z_0. However, the retinal size ratio cannot signal metric depth between two objects: like relative disparity, when no other cues to distance are available it can only convey relative distances but not absolute distance.

To combine depth from the retinal size ratio with the disparity-defined depth, we first generate a probability density function for the retinal size ratio. First, we assume that the visual system interprets the retinal size information of the two targets as being generated by two objects of equal size (e.g. the relative size cue, as it is commonly known). Second, we assume that the measurement of the retinal size ratio between the two targets is Gaussian in the log-domain, similar to how we generated the PDFs for disparity and vergence. This skewed distribution is used as an approximation of the ratio distribution, which would be the result of the ratio of two retinal size measurements, assuming that the each measurement of retinal size is normally distributed. Finally we assume that the PDF for the retinal size ratio has a standard deviation that is a factor of four higher than that from relative disparity and vergence. The reasoning for the latter assumption is as follows. From experiment 1's "different-size" condition, we know that the average perceptual bias is about 2,5 cm for a retinal size ratio of 2 at an absolute distance of 50 cm. However, in the case of a retinal size ratio of 2 the depth estimate specified by retinal size would be 50 cm (one object would be seen at a distance of 50 cm and the other would be seen at a distance of 100 cm). Yet, the much lower value for the perceptual bias (about 5% of the depth specified by the retinal size ratio) suggests that retinal size has a relatively weak influence compared with the disparity-defined depth estimate (i.e. the standard deviation for the relative distance ratio from retinal size should be much higher, relative to the disparity). We chose a value that was fourfold the

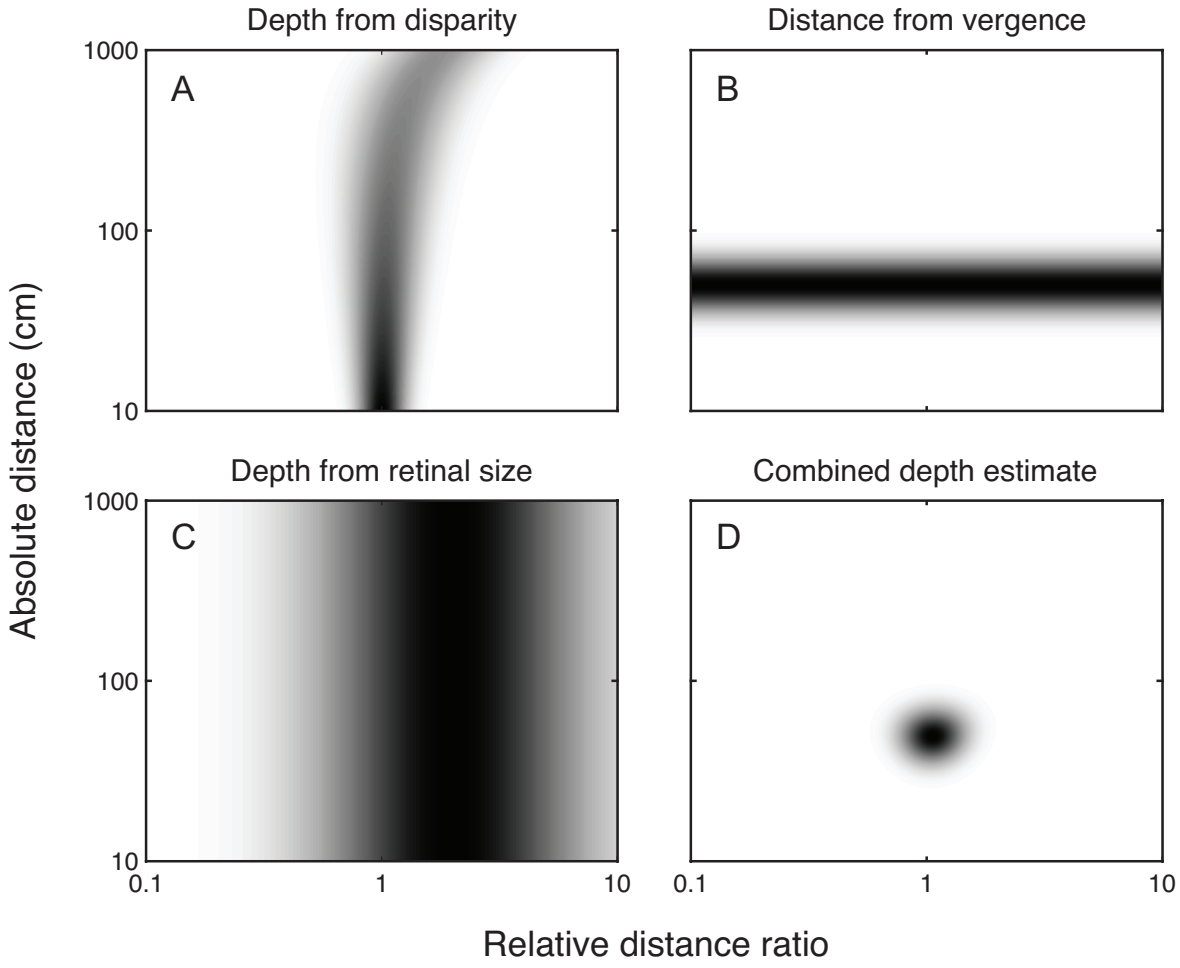

Figure 5.12: Probabilistic model of the combination of disparity, vergence and retinal size information (A) the probability distribution $P(z_0, r|\delta)$, where δ is the measured relative disparity. (B) the probability distribution $P(z_0, r|\gamma)$, where γ is the measured vergence distance. (C) the probability distribution $P(z_0, r|\theta)$, where γ is the measured retinal size ratio. (D) posterior distribution $P(z_0, r|\delta, \gamma, \theta)$ which is the product of the distributions in (A), (B) and (C). From this distribution the absolute and relative distances can be estimated.

standard deviation of disparity and vergence so that the model provided reasonable fits to our psychophysical data from Experiment 1.

Figure 5.12C shows the probability density function for depth from a retinal size ratio of two. The most likely value of relative distance ratio specified by this function (the mean value) specifies a straight line through a relative distance ratio of 2: under the assumption of equal size, the relative size ratio can specify a relative distance, but this relative distance can be generated by two objects at any absolute distance. The wideness of the function relative

to that for disparity reflects the increase in the standard deviation of the relative depth estimate from relative size. Again, we apply Bayes' rule to combine the information from relative disparity, vergence and retinal size. In essence, we multiply all three probability density functions and we use the Maximum Likelihood Estimate (the most likely value of the absolute and relative distance) to obtain the relative and absolute distance that are most likely given the visual information to calculate the predicted depth. Figure 5.12 provides an illustration of the combination process. The combined depth estimate (Figure 5.12D) is the product of the depth-from-disparity PDF (Figure 5.12A), the distance-from-vergence PDF (Figure 5.12B) and the depth-from-retinal-size PDF (Figure 5.12C). Here, the inclusion of the relative retinal size information produces a shift in the maximum probability towards larger values of relative distance ratios and, as a result, a perceptual bias with respect to the true disparity-defined depth (this subtle shift is not directly evident from the figure).

5.7.3 Predicting perceptual bias

Having constructed a basic model, we can now test how well this model predicts the perceptual biases that we found in Experiments 1 to 3. We define the predicted shift in the PSE from the model as follows. We assume that the shift in the PSE specifies the perceived depth between the two targets when, in fact, they are at the same distance (i.e. when the depth and thus the disparity between the two objects is zero). However, we cannot specify the disparity-defined depth for a disparity value of zero (because we cannot divide a scalar by zero; see equation 1.4). To circumvent this problem, we calculated model predictions over a range of disparity values, symmetrically spaced around zero (±26 arcmin, with intermediate steps of 4). The average of the predicted perceived depths calculated in this fashion should then accurately specify the shift in the PSE when there is no depth between the two targets. Because the depth

that is predicted from disparity and vergence without retinal size information is equal to the true depth specified by geometry, we specify the predicted shift in the PSE as the difference between the PSE predicted from disparity and vergence information without retinal size from the PSE predicted from disparity, vergence and retinal size.

Experiment 1 We first investigate how well the model predicts the PSE shifts from Experiment 1's "different-size" condition. Figure 5.13A shows the predictions of the model over a range of disparities, using a size ratio of two. To calculate the PSE, we first calculate the model without retinal size information, which provides the (baseline) unbiased disparity-defined distance. We then calculate the model again but include the retinal size information - the difference between the PSE's is what defines the shift in the PSE. Consistent with our psychophysical results, in our model a difference in the retinal size information between two objects results in a shift in the perceived depth of about 2.7 cm. What if two objects are of equal retinal size? When we set the retinal size ratio to 1, we find that consistent with Experiment 1 ("same-size" condition) the combination of disparity-defined depth and retinal size does not result in systematic perceptual bias (Figure 5.13B). There is a slight gradient in the shift in the PSE, which seems proportional to the difference between the relative distance ratios of the disparity-defined depth and the retinal size ratio. As this difference becomes smaller, so does the influence of the relative contribution of the retinal size information, but the average shift in the PSE is equal to zero.

Experiment 2 Thus far, our model predicts the pattern of results in Experiment 1 reasonably accurately. In Experiment 2, we found that increasing the retinal size ratio resulted in an increase in the perceptual bias; perceptual bias remained constant when we kept the retinal

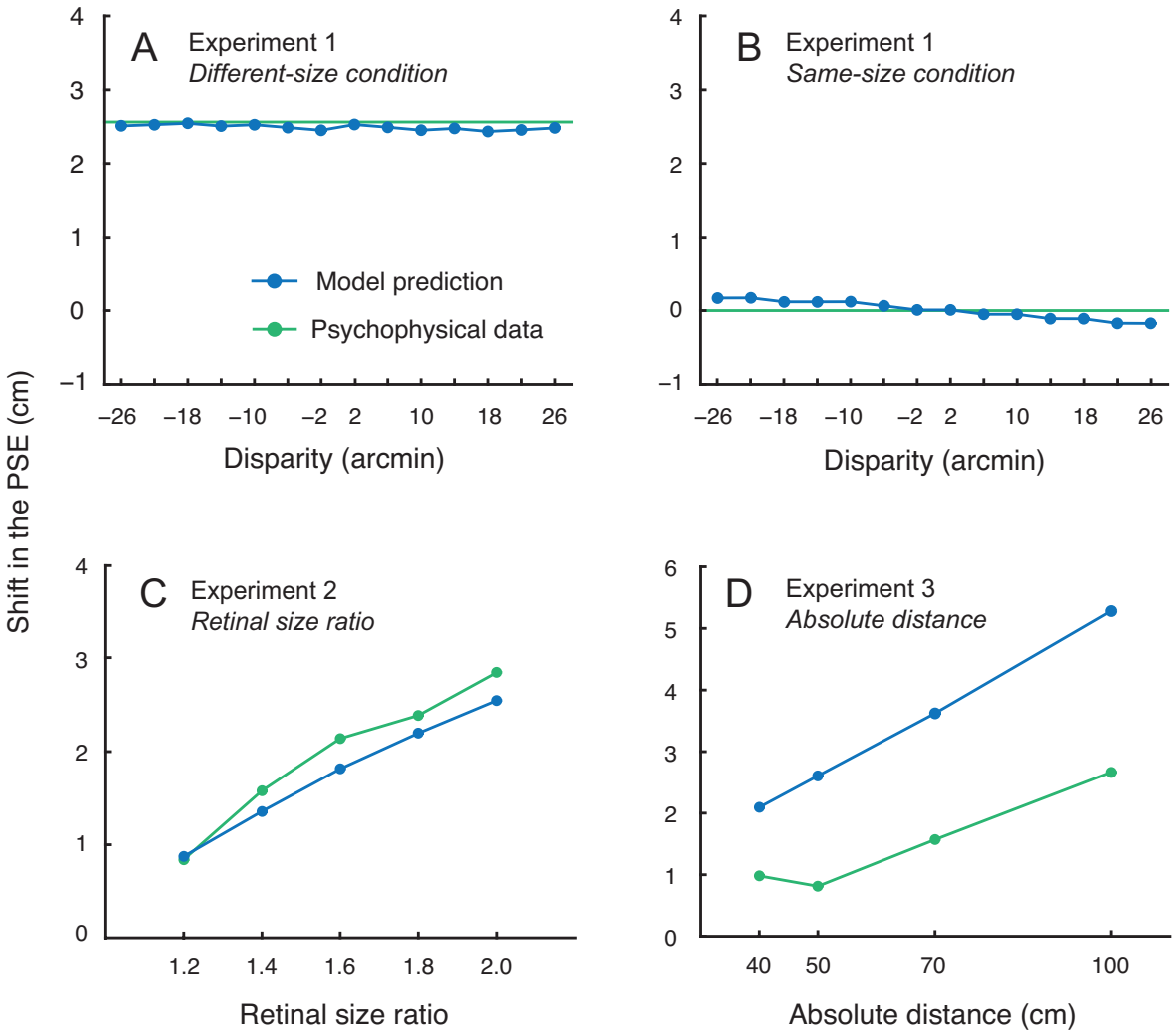

Figure 5.13: Model predictions (blue data series) and psychophysical results (green data series) for experiments 1 - 3. (A) Predictions for Experiment 1's different-size condition. (B) Predictions for Experiment 1's same-size condition. (C) Predictions for Experiment 2, in which we manipulated the size ratio between two targets. (D) Predictions for Experiment 3, in which we measured bias at different absolute distances.

size ratio constant but increased the absolute sizes of the two targets. We can explain this pattern of results using our model: varying the mean of the retinal size ratio PDF results in a shift of the PDF on the relative distance ratio axis (i.e. a horizontal shift). Systematically varying the mean of the retinal size ratio PDF in our model like we did in Experiment 2 results in a proportional change in perceptual bias (Figure 5.13C). This change is approximately linear and can account for an increase in perceptual bias as a function of an increase of the retinal size ratio, found in Experiment 2.

Experiment 3 In Experiment 3, we found that an increase in the viewing distance increased perceptual bias. This may reflect an increase in the uncertainty in the estimate of relative depth from relative disparity as the viewing distance increases, an increase of the uncertainty of the vergence distance, or both. Similar to a weighted average approach, the uncertainty in the individual estimates determines the relative influence they exert on the final estimate. A cue that provides a low uncertainty (i.e. a low standard deviation) will have a large influence; conversely, a cue with high uncertainty (i.e. a high standard deviation) will have a small influence. As absolute distance increases, the increased uncertainty in the disparity-defined depth estimate reduces the impact of the disparity-defined cue and increases the influence of the retinal size cue. We can test this idea in our model by increasing the mean of the distance-from-vergence PDF. Systematically increasing the absolute distance in our model, we find an approximately linear increase in the amount of perceptual bias, consistent with the pattern of psychophysical results of Experiment 3 (Figure 5.13D).

Taken together, the predictions from our model account reasonably well for the general pattern found in the psychophysical data: it predicts no bias when there is no retinal size difference, but as we introduce a retinal size difference, our model predicts a shift in the PSE (Experiment 1). In addition, there was a proportional change in the PSE shift when we varied the retinal size ratio (Experiment 2) and the absolute distance (Experiment 3). However, for Experiment 2 and especially for Experiment 3 we find large discrepancies between the psychophysical data and the model predictions and therefore the model does not provide a good fit to these data. The most likely reason for this inaccuracy of the fit is that we largely used different observers for each experiment. To model our predictions, we used the same values of standard deviation across all experiments. To illustrate, in our model predictions the PSE shift at 50 cm and a retinal size ratio of 2 predicted for Experiments 2 and 3 is equal

to the shift we predicted in the "different-size" condition in Experiment 1. However, our results from Experiment 1 show that there were large differences in the amount of bias that was shown by different observers and it is highly likely that different groups of observers will - on average - produce average PSE shifts that differ. This discrepancy poses a limitation to the predictive power of the model in its current state. Better model fits would be expected across experiments if the same observers participated in all experiments conducted in this chapter.

As outlined before, the exact values for the standard deviations that were chosen to implement the model are not directly motivated by our psychophysical data, but instead were chosen mainly to fit our model predictions to the results of Experiment 1. Although the model seems to account for the qualitative aspects of our data fairly well, the fact that the standard deviations are free parameters poses an important limitation, as the same predictions of bias can potentially be produced by multiple combinations of different standard deviations for the different sources of information. To constrain the model, and potentially provide a better quantitative account of our results, these free parameters would ideally be known. Specifically, to the quantitative fits of the model we would have to know observers sensitivity to (1) estimating absolute distance from extra-retinal cues to vergence and (2) judging relative depth from the relative size ratio. Taking into account the large variations in thresholds between observers (and the proportional amount of perceptual bias), the model can then be fitted to individual observers' results, rather than on average data.

In summary, in three experiments we have investigated how retinal size information interacts with relative disparity information and the vergence orientation of the eyes. We find that a target projecting a larger retinal image size is seen as closer than a target projecting a smaller image size at the same disparity-defined distance. In addition, we show that this perceptual

bias changes as a function of (1) the ratio of retinal sizes but not absolute size differences; and (2) the viewing distance. To account for these findings, we first explored a Bayesian cue-combination model that combines relative disparity with the absolute distance obtained from the vergence angles of our eyes. We then showed that a Bayesian model, in which to interpret the retinal size information, the visual system assumes that the retinal projections of the two targets presented on each trial were of generated by two objects of equal physical size, even though in reality they were not. It is this equal-size assumption that causes the perceptual bias in the combination with relative disparity and vergence.

These findings have implications for Bayesian models of cue combination. Traditionally, Bayesian cue combination frameworks are described as optimal, where the optimality criterion is defined as reducing the variance of the final (cue-combined) estimate. This is a sensible approach to take when cues specify the same depth (i.e. they are unbiased) because the final estimate would remain unbiased. However, when there is a cue-conflict between cues, the visual system may operate in a more robust fashion and 'veto' (discard) the biased cue (Landy et al., 1995; Blake, Bulthoff, & Sheinberg, 1993; Banks & Backus, 1998). In our experiment, the information conveyed by retinal size and disparity is - in principle - unbiased; the targets are of different size and this leads to differences in the retinal projection. However, in interpreting the retinal size information, the visual system seems to assume that the two targets are of equal size. This assumption then introduces a large bias in the estimate of depth obtained from retinal size cue with respect to the relative disparity cue, a bias that is proportional to the ratio of retinal sizes; yet, our results suggest that the visual system does not veto the biased cue, even though disparity provided a reliable estimate of depth. Conversely, our results show that the optimisation of cue combination with respect to just the variance (Hillis et al., 2004, 2002) may result in a perceptual bias in the cue-combined estimate.

Our results suggest that the visual system utilises *all* these sources of information to judge the depth between two unfamiliar objects, even though this leads to systematic inaccuracies in judgments of depth. The Bayesian model that was presented as part of this thesis behaves in an approximately similar manner as our observers and provides a reasonable account for the qualitative trends that we found in our data. However, there remains a discrepancy between the amount of bias predicted by the model and the measured bias. One reason for this discrepancy may be that we used different observers for each experiment. The idiosyncrasy of the perceptual bias - evident from the results of Experiment 1 - compliates fitting the model across different experiments. Future work will attempt to improve the *quantative* model fits and reduce the models' free parameters.

6

Physical object size affects judgments of three-dimensional speed

In the previous chapter, we found that differences between the retinal size of two objects systematically affect the perceived depth between them: a larger object was seen as closer than a smaller target even though they were at the same disparity-defined distance. This chapter investigates whether this is also true for motion-in-depth. To test this, we sequentially presented two approaching targets of (1) equal physical size and (ii) different physical size over the same range of distances. Our results indicate that object size systematically affected judgments of 3D speed, so that a larger target was reported as approaching faster than a smaller target moving over the same range of distances. When two targets were of equal size, there was no perceptual bias. We show that this result cannot be explained exclusively by a difference in the perceived distance of the two objects, as caused by differences between their retinal sizes. Rather, we suggest that the difference in distance from retinal size interacts with differences in the looming signal.

6.1 Introduction

A key function of the visual system is to provide information about objects moving in depth, so that we can initiate an interceptive or evasive actions when we need to. In natural viewing, motion in depth (3D motion) can be signalled by a variety of sources of perceptual information, such as changes in the relative disparity over time (the "Changing Disparity" cue: Cumming & Parker, 1994; Harris & Watamaniuk, 1995), differences in the interocular velocity (Beverley & Regan, 1973; Rokers et al., 2009; Shioiri et al., 2000) or their combination (Harris, Nefs, & Grafton, 2008). In addition to binocular cues, compelling illusions of 3D motion are produced by monocular cues, such as gradual changes in the size of an object's retinal image over time (looming). Over 150 years ago Wheatstone (1852) showed that an isotropic expansion of an object's retinal image creates the impression that the object is approaching, even though the object is in fact stationary. This suggests that, in interpreting looming information, the visual system seems to resolve this initial ambiguity by assuming that objects are rigid and thus favors an interpretation of 3D motion (i.e. motion towards or away from the observer) over the geometrically correct alternative possibility that the object's physical size changed.

Previous work on 3D motion has demonstrated the importance of both looming and disparity in perceiving 3D motion. For example, when vergence demand of a large stimulus (i.e. the absolute disparity of the entire stimulus) is modulated in the absence of looming information, observers perceive no 3D motion: the stimulus is perceived as static (Erkelens & Collewijn, 1985b; Brenner et al., 1996; Welchman et al., 2009; Regan et al., 1986) even though observers make vergence pursuit eye movements with high gain (Erkelens & Collewijn, 1985a). The cause of this effect has been shown to be (1) the conflict between changing vergence signals and the absence of looming (Howard, 2008; Gonzalez et al., 2010; Brenner et al., 1996);

and (2) the absence of a changing relative disparity signal (Erkelens & Collewijn, 1985b; Brenner et al., 1996). Not surprisingly, Brenner et al., 1996 showed that judgments of 3D speed were most accurate when both looming and disparity cues signaled 3D motion.

A closely related topic in 3D motion in which the interaction between binocular and monocular cues has been investigated extensively is time-to-contact (i.e. the time remaining until an object hits us or an external point). Early theoretical accounts of time-to-contact (TTC) estimation have been largely dominated by the so-called 'tau-hypothesis', which maintains that a perceptual quantity called 'tau' (the ratio of an object's retinal size to its looming rate) is directly available from the monocular retinal image and provides the information necessary for interceptive timing (e.g. Lee, 1976; Lee & Reddish, 1981; Lee et al., 1983; Regan & Hamstra, 1993; Todd, 1981, also see Chapter 3 of this thesis). An attractive feature of the tau-hypothesis lies in the fact that - theoretically - it should be independent of the object's distance and physical size. As a result, purely monocular information can theoretically provide a direct temporal estimate of TTC. However, other studies have shown that these early accounts underestimated the role of binocular information in judging TTC (e.g. Heuer, 1993; Gray & Regan, 1998; Rushton & Wann, 1999) and that when both looming and changing disparity information are available, the visual system may use both. Specifically, there is strong evidence to suggest that observers make more accurate estimates of TTC by combining the two sources of information according to either the reliability of individual cues (Gray & Regan, 1998) or by giving the highest weight to the cue specifying a more imminent TTC (Rushton & Wann, 1999).

The aim of this chapter is to further investigate how binocular disparity interacts with looming information in judging 3D motion. In particular, we are interested in how an object's physical size affects observers' judgments of 3D speed. We next consider the geometry of

the available information. The binocular correlate of 3D speed (v_δ) is traditionally described (Regan, 2002) as a function of the inter-ocular separation (i), the rate of change in disparity ($d\delta/dt$) and the viewing distance (z) squared:

$$v_\delta = \frac{z^2(d\delta/dt)}{i} \tag{6.1}$$

Thus, when two objects of differing size travel the same trajectory and binocular cues are available, disparity in isolation specifies the same 3D "disparity-defined" speed for both objects; the estimate of 3D speed should be unaffected by physical size. In contrast, the 3D speed of an approaching object is described as a function of its looming rate ($\dot{\theta}$), its retinal size (θ) and, most important, its physical size S (Lopez-Moliner, Field, & Wann, 2007):

$$v_\theta = \frac{\dot{\theta}}{s\theta^2} \tag{6.2}$$

Alternatively, under small-angle approximation (Regan & Beverley, 1979), 3D speed can be specified as a function of the looming rate, the physical size of the object and its viewing distance squared:

$$v_\theta = \frac{\dot{\theta}z^2}{s} \tag{6.3}$$

Figure 6.1 shows how different object sizes result in different looming rates. It follows that any measurement of looming cannot directly specify an object's 3D speed without knowledge of the physical size of the object. In other words, a single looming rate can be consistent with infinite combinations of physical object sizes, approach speeds, and physical distances (note that in Eqn 6.2 an object's retinal size also depends on its distance from the observer). Thus

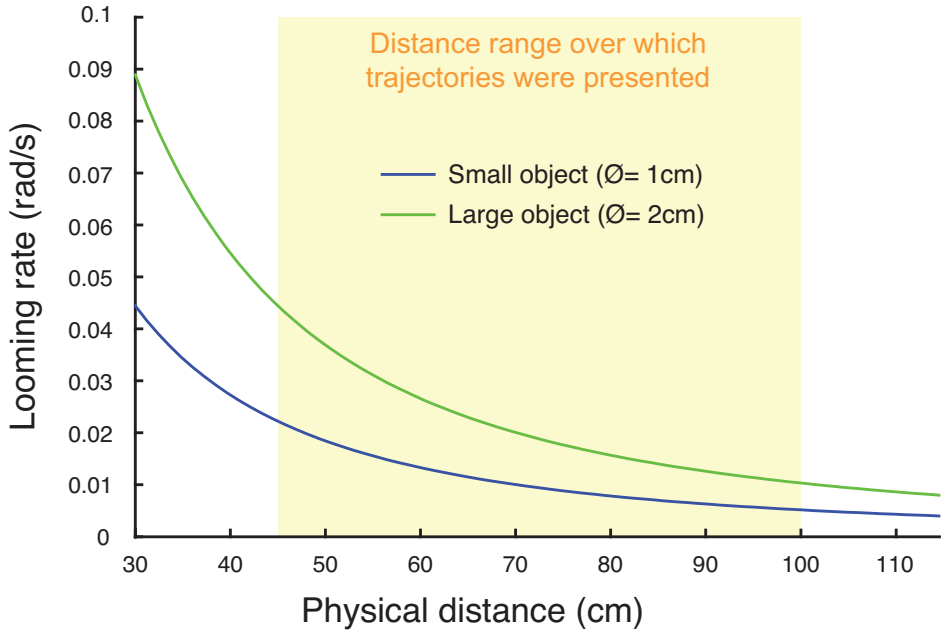

Figure 6.1: Differences between the looming signals of two objects of different physical size. The green curve shows the looming rate of a larger object (diameter = 2 cm); the blue curve shows the looming rate of a smaller target (diameter = 1 cm). The yellow area shows the range of distances over which the target moved towards the observers.

looming rate in isolation should not support accurate judgments of 3D speed: the looming signal is ambiguous with respect to the 3D speed of an approaching target.

We propose that, due to this ambiguity of the looming signal, differences in an object's looming rate (caused by a difference in the physical size) should not affect judgments of 3D speed when both looming and disparity information is available. Rather, in close proximity the disparity signal should be highly reliable and observers should be able to judge 3D speed accurately on the basis of this information alone. To test this idea, we measured psychometric functions of "disparity-defined" approach speed (i.e. the 3D speed signaled from the combination of changes in vergence and changes in relative disparity). On each trial, observers sequentially viewed the motion excursions of two approaching targets. In one condition, observers saw two targets of equal size (i.e. two small targets or two large targets) at the same distance. In another condition observers saw two targets of differing size (i.e. one small and

one large target). We find that, even though the disparity-defined 3D speed specified the 3D speed accurately, observers' judgments of 3D speed were significantly affected by target size.

6.2 Methods

6.2.1 Observers

Nine observers were recruited from staff and students from the University of Birmingham, all of whom were naive to the purposes of the experiment. All had normal or corrected-to-normal vision and gave informed consent. Prior to participation all observers were screened to ensure that they could discriminate 1 arcmin of disparity in a briefly (300 ms) presented random dot stereogram.

6.2.2 Apparatus

Stimuli were presented stereoscopically in a dark laboratory using a two-monitor haploscope in which the two eyes viewed separate 21 inch CRTs (ViewSonic FB2100x) through front-silvered mirrors. Viewing distance was 50 cm. We adjusted the haploscope so that the inter-ocular separation and the vergence angle were configured correctly for each individual observer. Stimulus presentation was controlled by a Windows PC with an NVIDIA Quadro FX4400 graphics card. The monitors displayed 1600 x 1200 pixels at 100 Hz. Individual pixels subtended about 1.75 x 1.75 arcmin. Head movements were restricted using a chin rest. Responses were recorded using the keyboard.

6.2.3 Stimuli

Figure 6.2 shows a cartoon of the stimulus used in this study. Targets were large (diameter=2cm) or small (diameter=1cm) outline discs, which approached at a constant speed along the line of sight. In addition to the target, we showed a background of textured cubes that was

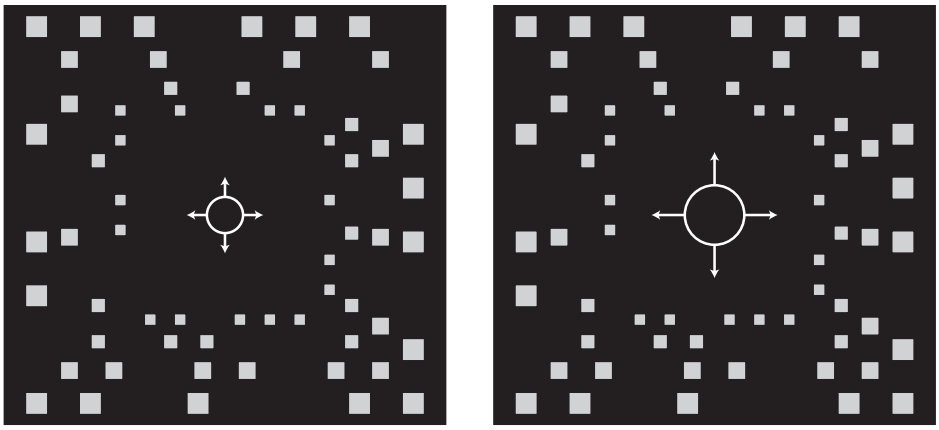

Figure 6.2: Frontal view of the stimulus configuration The left panel shows a small target; the right panel shows a large target. The target (the outline disc) was surrounded by a background of textured cubes (here represented by untextured squares) that receded in depth. The squares formed a tunnel around the maximum motion excursion of the target. The arrows around the target represent a looming signal. Note that in the actual experiment, the targets were shown binocularly and disparity signals to 3D motion were available in addition to the looming signal.

configured as a square tunnel around the motion trajectory of the target. This tunnel was visible throughout the entire experiment and extended from 40 cm to 105 cm from the observer's eyes. The distance from the centre of the target to the edge of the background was about 10 cm. Stimuli were created using C# and OpenGL graphics libraries and were rendered using anti-aliasing and geometric perspective projections from each eye, taking the observer's interocular separation into account. Using perspective projection ensured that the targets' retinal size changed in size in accordance with its disparity-defined distance, such that there was no cue conflict.

6.2.4 Procedure

Observers (the author and seven naive observers) sequentially viewed two motion intervals, each containing an approaching target, and judged whether the target approached faster in the first or second interval. We measured speed discrimination thresholds in four conditions. In the first two condition observers saw one large and one small approaching target (different

size condition); in the other two conditions they saw two large or two small approaching targets (same size condition). This resulted in four combinations of object size; Large-Small, Small-Large, Large-Large and Small-Small, where the first is the reference size. To ensure that observers did not know the position of the reference and test, we presented them in a quasi-randomised order on each trial. The reference stimulus always approached the observer at 40 cm/s. The approach speed of the test stimulus was varied in seven steps between 26 and 54 cm around the mean speed of 40 cm/s. On each interval, the target appeared at its starting distance and remained there for 500 ms. It then started approaching the observer along the line of sight and stopped at its end distance. To ensure that observers could not judge velocity on the basis of presentation duration or distances, we randomised the start distance (90±10cm) and end distance (55±10cm) of each interval. The shortest motion excursion could be 15 cm; the longest 55 cm. The minimum presentation duration was about 300 ms (15 cm at 54 cm/s); the longest presentation duration was about 2 seconds (55 cm at 26 cm/s). Each observer completed 20 repetitions for each unique combination of stimulus value and condition (560 trials in total) in two separate sessions.

6.3 Results

6.3.1 Data analysis

To quantify our results we first constructed average psychometric functions for each condition (collapsed across observers because individual data for each condition did not differ significantly) by plotting the proportion of responses in which the observer reported that the test stimulus approached faster as a function of the simulated speed of the target. We fitted cumulative Gaussians to these data using the bootstrapping technique implemented by the psignifit toolbox (Wichmann & Hill, 2001). The 95% confidence intervals were obtained us-

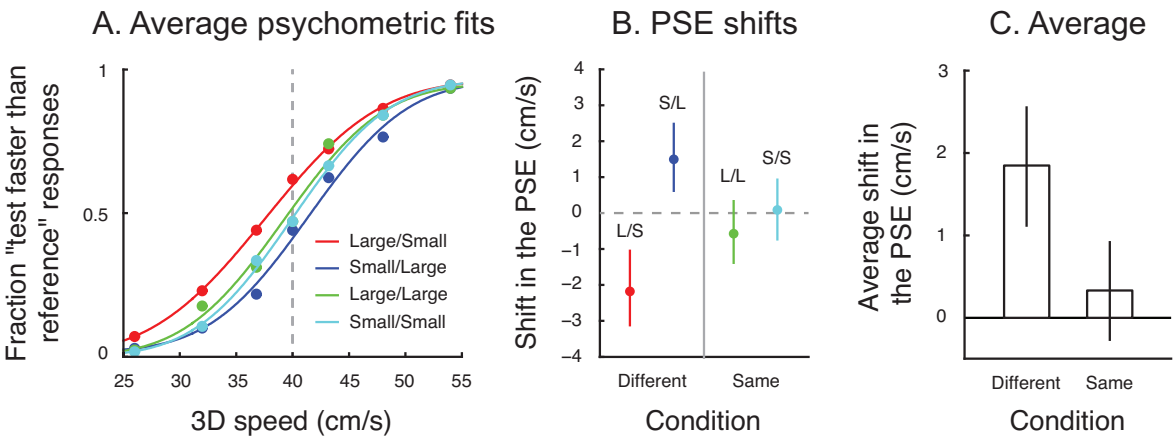

Figure 6.3: summary of the psychophysical results. (A) average psychometric fits for different sized targets (blue and red data series) and same-sized targets (green and cyan data series). The red data shows the average results for trials in which the large target was the test stimulus and the small target was the reference stimulus. The blue data shows the average results for trials in which the small target was the test stimulus and the large target was the reference stimulus. Green data shows trials in which two large objects were presented; cyan data shows trials in which two small targets were presented. (B) the shift in the PSE's for different sized targets (blue and red data) and same sized targets (green and cyan data). Error bars show 95% confidence intervals obtained from a bootstrapping method. (C) Average shift in the PSE for different-sized and same-sized targets. See text for more information on how this average was calculated.

ing the percentile method and were based on 1999 simulations. We quantified the point of subjective equality (PSE) as the mean of the fitted Gaussian (i.e. its position on the x-axis) and the discrimination threshold as its standard deviation (i.e. the slope parameter) divided by $\sqrt{2}$. To directly quantify the difference between the perceived speed of large and small targets, we averaged the unsigned bootstrapping results for the PSE's of the two different-size (large-small and small-large) and the two same-size (large-large and small-small) conditions, resulting in two sets of 1999 average bootstrap results; one set of 1999 average bootstraps for the same-size condition, the other for the different-size condition). We then used these averaged data to calculate the average shift in the PSE's by calculating the median (the actual PSE) and its 95% confidence intervals.

6.3.2 Perceived three-dimensional speed

Our results show that PSE's were different across conditions ($F_{1.26,10.16} = 5.56, p < .05$; corrected for violation of sphericity). Specifically, there was no shift in the PSE when the two objects were of equal physical size (Figure 6.3A, green vs. cyan data series; $t_8 = -.65, p = .53$). We found a slight shift in the PSE's for different sized objects (Figure 6.3A, blue vs. red data series; $t_8 = -2.47, p < .05$). Specifically, relative to a large reference target, a small test target was seen as approaching about 2.5 cm/s slower; relative to a small reference target, a large test target was reported as approaching about 2.25 cm/s faster (Figure 6.3B). On average, the shift in the PSE between large and small targets was about 2.3 cm/s (Figure 6.3C). Although short of significance ($F_{1,9} = 0.4, p = .94$) it is clear that there is substantial variation between observers' PSE's (see Figure 6.4): some observers hardly show any shift in the PSE (e.g. subjects 2, 3 and 7) whereas others display relatively large shifts of up to 10 cm/s (e.g. subject 8). Thresholds differed between observers ($F_{1,8} = 246.98, p < .01$), but there was no significant difference between conditions ($F_{3,24} = .51, p = .68$).

6.4 Discussion

Judging 3D motion from two-dimensional images is a difficult problem, as the monocular input to the two eyes is inherently ambiguous. When an object's size is unknown, it seems inefficient for the visual system to rely on ambiguous looming information when 3D speed is accurately signaled by changing disparity. Yet, our results show that differences in physical size subtly but systematically affect judgments of 3D speed, such that larger objects are reported as approaching faster than a smaller object approaching at the same 3D speed. Like the results from the previous chapter, these results are somewhat surprising: strictly speaking, both the looming and disparity information were generated in a geometrically correct man-

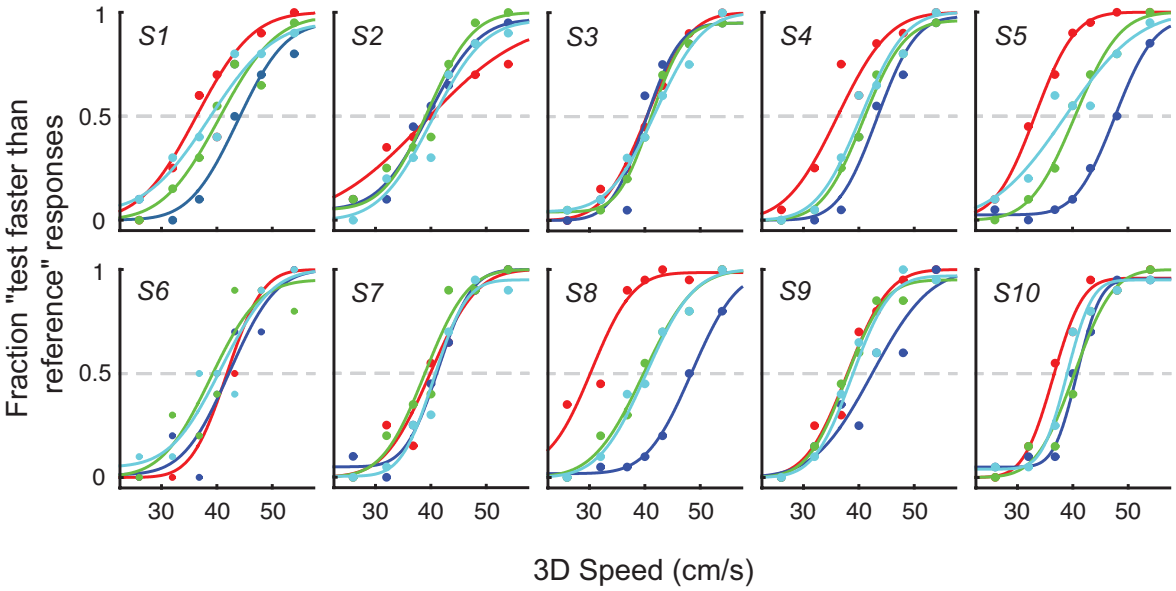

Figure 6.4: Individual observer data for ten observers. The red data shows the average results for trials in which the large target was the test stimulus and the small target was the reference stimulus. The blue data shows the average results for trials in which the small target was the test stimulus and the large target was the reference stimulus. Green data shows trials in which two large objects were presented; cyan data shows trials in which two small targets were presented.

ner (i.e. there was no cue conflict). The only difference between the two targets lies in their physical size, and thus their relative looming rates and retinal sizes. In the introduction we argued that that differences between the looming rates of two objects - in principle - should not inform judgments of 3D speed. Yet, the looming information is not ignored, as was initially anticipated. The question - then - is how the combination of changing disparity and looming results in a perceptual bias. We explore this in the next section.

6.4.1 How does object size affect judgments of 3D speed?

Perhaps the most obvious explanation of our results is that the visual system takes a weighted average (Landy et al., 1995) of the 3D speed estimates derived from the looming signal and that derived from changing disparity signals; the small magnitude of the perceptual bias may reflect a low weight of the estimate of 3D speed from looming, relative to that from changes in disparity. There is good evidence that shows that optimal cue combination occurs in motion-

in-depth (Chapter 4, this thesis; Rushton & Wann, 1999). However, there may be a different interpretation. In the previous chapter, we found that the retinal size ratio between two objects systematically affected their perceived relative distance, assuming that the retinal projections of the targets were generated by two objects of equal size. In particular, a target with a larger retinal size was seen as closer than a target with a smaller retinal size at the same distance. In the Introduction, we showed that the 3D speed estimate from disparity can be computed independently from knowing the target's physical size (see Eqn. 6.1). However, the recovery of 3D speed from disparity depends on an estimate of the viewing distance. Similar to the experiments in Chapter 5, there was an inherent difference in the retinal size between two targets from differences in physical size. As a result, it is likely that the perceived distances of small and large targets also differed. If this biased perceived distance is then used to scale the changing disparity signal into an estimate of 3D speed, this might account for our results.

To explore whether this approach accounts for our data, we first assume that - consistent with the results from Chapter 5 - a target with a larger retinal size is seen as closer than a target with a smaller retinal size. Considering the geometry of binocular vision (Figure 6.5) we can see that a given value of disparity gives rise to a larger displacement when the absolute distance as the absolute distance increases. In other words, of a target that is seen as farther should be perceived as approaching at a faster rate (because the displacement over time is larger) and, conversely, a target that is seen as closer should be perceived as approaching at a slower rate (because the displacement over time is smaller). This is opposite to our results, as the larger target was reported to approach at a faster rate. We therefore conclude that a bias in the perceived distance from differences in the retinal size cannot account for our results, though this does not mean that differences in the perceived distance of small and large targets do not affect the 3D speed from disparity. However, it is clear that these differences cannot provide a

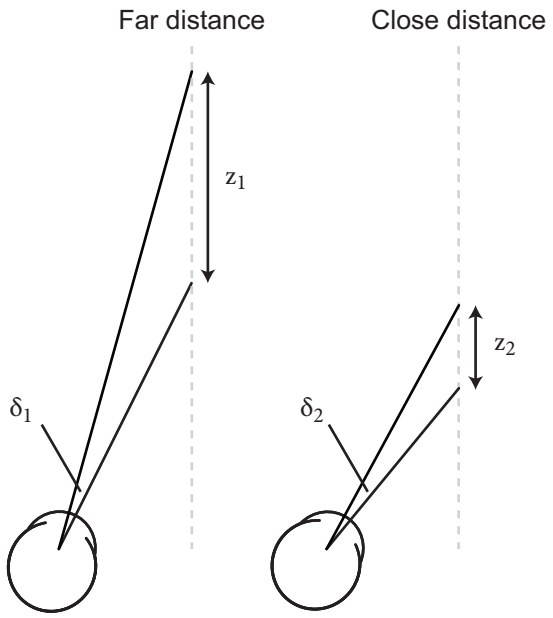

Figure 6.5: The same value of disparity results in different extents of depth at different absolute distances. Two equal values of disparity (here shown as δ_1 and δ_2) result in different depths (z_1 and z_2) at different absolute distances.

full explanation. Rather, we suggest that the perceptual bias found in this experiment is most likely caused by an interaction between (1) differences between the looming rates of the two targets; and (2) their biased perceived distances, as caused by differences in retinal size.

One potential concern to the interpretation of our results is due to observers' eye movements - which we did not measure - and the weighting of the binocular information. If observers kept their vergence fixed, the target would have moved over the retinas in opposite directions in each eye. This changing disparity signal with respect to fixation would have provided a strong disparity signal to judge the velocity of 3D motion. However, it is highly likely that observers pursued the target as it approached. Assuming that vergence pursuit was fairly accurate, the image of the target on the retinal would be relatively stable (i.e. its absolute disparity would remain approximately constant at zero) and estimates would mainly rely on estimates of vergence pursuit from extra-retinal information. As has been discussed in the introduction of Chapter 4, it is well established that observers' ability to perceive 3D motion is severely reduced when the changes in the vergence angles of the eyes are not accompanied by changes in relative disparity of static structures that provide retinal flow information

to estimate vergence pursuit (e.g. Brenner et al., 1996; Erkelens & Collewijn, 1985b; Harris, 2006; Regan et al., 1986). This could potentially severely reduce the contribution (or weighting) of the binocular signals with respect to the monocular looming signal. However, in our experiment we provided a static background; as the observers pursued the target, the static background provided retinal flow information to vergence pursuit through changes in the relative disparity of the background. This information could then be used to accurately judge the velocity of 3D motion from disparity.

The static background may be crucial not only in recovering 3D motion from disparity signals, it may also play a role in determining the reliability of the cue. Specifically, we find a slightly lower bias for larger objects. This may reflect a difference in the reliability between the disparity information generated by the large and the small targets; the edge of the large target is physically closer to the background (i.e. the eccentricity is smaller), thereby making it easier to extract the relative disparity between the two. As a result, the disparity information associated with the large target may be more reliable and when looming and disparity are optimally combined we would see a lower weight assigned to the looming cue.

6.4.2 Relation to previous studies

It is difficult to draw comparisons in relation to previous work on 3D speed. Most of the work on 3D motion has involved measurements of precision rather than accuracy (cf. Rushton & Duke, 2009). Indeed, few studies have shown an effect of size on 3D motion. Notable exceptions are studies by DeLucia (1991, 2005), in which an effect of object size was found when looming alone specified motion-in-depth (DeLucia, 1991). In particular, DeLucia (1991) monocularly presented two approaching squares of different physical size and asked observers to judge which square would hit the observer first. Relative size (assuming that the two ob-

jects were of equal physical size) indicated that the larger object was closer, but looming was manipulated so that 'tau' specified that the smaller object would arrive first. She found that the larger object was favored in judging which object arrived first, an effect known as the 'size-arrival effect' (DeLucia, 1991; DeLucia & Warren, 1994). Yet, when both disparity and looming specified 3D motion no effect of size was found (DeLucia, 2005). However, there is an important difference between these studies and ours; whereas DeLucia presented the two objects side by side, we presented the objects sequentially. It is well known that sensitivity to relative disparity (i.e. in the case of simultaneous presentation) is about tenfold higher than when the visual system has to rely on two measurements of absolute disparity as is the case of sequential presentation (Westheimer, 1979). This may also be the case with 3D motion; when objects are presented side-by-side a direct comparison between the two objects is possible. If we assume that the visual system optimally combines information from disparity and looming, it is possible that the lack of the 'size arrival effect' in (DeLucia, 2005) reflects a high sensitivity to a difference in the rate of change in relative disparity between the two approaching objects. As a result, the weight assigned to looming may be so low that its effect cannot be discerned. In this sense, our sequential presentation method provides a more sensitive measure to investigate the effect of size.

6.4.3 Conclusions

Our results show that object size systematically affects judgments of 3D speed, so that a larger target is seen as approaching at a faster speed than a smaller target moving at the same speed and over the same range of distances. Our results cannot be explained by incorporating the results from the previous chapter. Specifically, a difference in the perceived distance of the two targets - as induced by differences in the retinal size between them - cannot account for our

data. Rather, we suggest that perceptual bias is most likely caused by an interaction between (1) differences between the looming rates of the two targets, perhaps coupled with assumptions about the physical sizes of the presented targets; and (2) their biased perceived distances, as caused by differences in retinal size (see Chapter 5). These interactions are potentially well-described by a Bayesian cue combination framework, but to implement this approach we first need to understand how the visual system computes three-dimensional speed from looming and disparity signals.

7

General discussion and conclusions

The work presented in this thesis applied psychophysical and computational techniques to gain further insight into two related questions: (1) which sources of perceptual information does the human visual system use to estimate depth and motion in depth; and (2) when multiple cues to depth are available, how does the visual system then combine these into a unified percept of depth? The general discussion is structured as follows. I first summarise the main findings from each experimental chapter. I then answer the main questions posed in this thesis. I close the thesis with a conclusion.

7.1 Summary of the main findings and contributions

7.1.1 Chapter 3

In the first experimental chapter we investigated which perceptual information the observers use to estimate time-to-contact (TTC) in two laboratory tasks: one absolute estimation task and one relative judgment task. In particular, we assessed whether perceptual judgments were based on the task relevant information (TTC), or whether observers based their judgments on one or more of its covariates (e.g. velocity or displacement). For both experiments we considered a wide range of covariates, including spatial, temporal and spatio-temporal variables. Importantly, the crucial manipulation was that the correlation between the covariates

was systematically varied over a range of randomisation levels. To determine the basis of observers' responses, we then determined how performance in both TTC tasks was affected by randomisation.

In the first experiment, we tested this using an absolute estimation task, in which observers pressed a button when they thought the target would have hit them. The results of this experiment showed that observers were severely inaccurate in their estimates, showing large systematic and idiosyncratic errors. To establish observers' basis of judgments, we conducted a principal components analysis (PCA) on all covariates and regressed the resulting components scores onto observers' estimated TTC - a statistical technique that to my knowledge has not been used before in this field. The results of this analysis did not present a unique solution: the component that best accounted for observers' judgments contained multiple variables, including TTC amongst several covariates. Although this provided some evidence that observers responded on the basis of TTC, the PCA technique was unable to fully dissociate TTC from other variables. However, observers' estimates of TTC were largely unaffected by randomisation - a result that to some extent indicates that judgments were based on TTC. However, explanations based on the use of other variables (e.g. the approach speed) rather than TTC could not be ruled out.

In the second experiment of this chapter, observers made TTC judgments with respect to an auditory cue (cf. Gray & Regan, 1998; Lopez-Moliner, Field, & Wann, 2007). The results of this experiment showed that observers' errors were now significantly reduced, most likely because the auditory cue provided a reference to 'anchor' their judgments to, thereby reducing the impact of systematic and variable errors that are associated with absolute estimation methods. To investigate which variable(s) provided the best account of observers' responses, psychometric functions for TTC judgments were first classified in terms of all the

covariates under consideration and fitted with Gaussians. We then compared the discrimination thresholds for TTC between individual covariates. These results indicated that only two variables could reasonably account for the data: the actual TTC and the time interval between the occlusion of the target and the auditory cue (a variable we called ΔT). To dissociate these two variables, we then showed that increasing he amount of randomisation (and thereby decreasing the correlation between these two variables) led to an increase in the discrimination threshold for δT, but those for TTC remained reasonably unaffected. This indicates that observers relied on TTC for their judgments.

The results from both experiments thus seem to confirm previous work (e.g. Regan & Hamstra, 1993; Gray & Regan, 1998), showing that observers do indeed judge TTC when asked to, rather than responding on the basis of covariates. However, there is a number of important differences between the work presented in this chapter and previous work, in particular with respect to the methods we used. First, previous work often created stimulus conditions in which a few variables were varied orthogonally in a factorial design (e.g. Regan & Hamstra, 1993; Regan & Vincent, 1995; Gray & Regan, 1998; Lopez-Moliner, Brenner, & Smeets, 2007). While this approach may seem attractive, this approach is unwieldy when more than two or three covariates are considered, leading to unrealistic large number of stimulus-condition pairings. Moreover, Regan and Hamstra (1993) used this factorial method to ensure that in their experiment the initial looming rate was orthogonal to tau and retinal size at the start of the trajectory. However, as the trajectory unfolds toward the observer this separation no longer holds. Finally, these "orthogonal" studies typically provide feedback, which complicates the interpretation of results in two ways. First, depending on the feedback regime, observers are able to discriminate covariates of TTC (e.g. the initial rate of expansion) with the same precision as TTC itself, making it difficult to know whether the task

reflects typical behaviour. Chapter 3 takes a different approach entirely, by varying the randomisation between TTC and its covariates. As we have shown, first classifying psychometric functions in terms of all the covariates under consideration and comparing how discrimination thresholds are affected by randomisation can then reveal the basis of their judgments, at least when using a relative task.

Providing feedback raises an additional concern: providing feedback generally increases the precision and accuracy of judgments. This will conflate the precision of judgments made in natural environments. In the experiments presented in this chapter no feedback was provided and we show that - in comparison with previous reports (e.g. Gray & Regan, 1998) - both the accuracy and precision of judgments are generally not as high. The reason for this may be that we used naive subjects, whereas many of the other studies on TTC - in particular the work of Regan and colleagues - often used a small number of experienced psychophysical observers (usually the authors themselves). This indicates that the view that humans exert exceptional precision in TTC tasks may have to be adjusted.

7.1.2 Chapter 4

This chapter investigated the contribution of extra-retinal signals to judgments of the speed of 3D motion (i.e. motion towards the observer). Until recently, it was widely held (Brenner et al., 1996; Erkelens & Collewijn, 1985b; Regan et al., 1986) that observers were insensitive to large changes in extra-retinal signals about the orientation of the eyes when they are not accompanied by changes in relative disparity (i.e. retinal slip). Specifically, previous work (e.g. Erkelens & Collewijn, 1985b; Brenner et al., 1996) showed that large changes in the absolute disparity of a large stimulus did not lead to sensations of 3D motion; As a result, it has been concluded that these signals provide poor information about motion in depth. More

recently, Welchman et al. (2009) have shown that extra-retinal information about changes in vergence is taken into account when judging the sign of motion (approaching vs. receding); such judgments are best explained on the basis that observers combine retinal slip with extra-retinal signals.

In this chapter we extended the technique developed by Welchman et al. (2009) so that we isolated extra-retinal signals to changes in the orientation in the eyes. Observers saw a small target, surrounded by a large background that moved laterally in opposite directions in each eye following a sinusoidal profile (consistent with approaching and receding motion). This - together with keeping the retinal size of the entire stimulus constant - induced unperceivable vergence pursuit eye movements. On each trial, the target's disparity relative to the background changed at different rates, so it was seen as approaching at different speeds. We then measured psychometric functions for 3D speed in four conditions: when the eyes were moving to (a) converge or (b) diverge and when the eyes were (c) maximally converged or (d) maximally diverged.

Our results showed that when the eyes were moving an approaching target was reported as faster during convergence and slower during divergence. In contrast, when the eyes were nearly static there was no difference in judgments. This demonstrates that extra-retinal signals support judgments of 3D motion magnitude as well as of sign. This is a new contribution, that has not been reported before. In addition, we showed that this result could not be the result of pursuit lag (i.e. lag in the vergence response with respect to the actual vergence demand) and we show that observers do not recover the actual real-world of the targets by scaling the angular measurements by the static eye vergence position. The latter finding confirms findings from previous work (e.g. McKee & Welch, 1989) that has shown that the visual system does not scale velocity in the fronto-parallel plane by eye vergence. Interestingly, this finding

suggests that the visual system codes 3D velocity signals in angular dimensions, rather than obtaining a real-world estimate.

To account for our findings, we then developed a model that optimally combines retinal signals (i.e. changes in the relative disparity) and extra-retinal signals (i.e. changes in the orientation of the eyes) to reproduce observers' perceptual judgments. Importantly, the model has no free parameters (i.e. we have used previous literature to derive a necessary parameter) and shows that although extra-retinal signals are taken into account, the contribution is relatively small. The results that we report in this chapter suggest that the contribution of extra-retinal signals - that was previously thought to be nil - needs to be qualified.

7.1.3 Chapter 5

This chapter investigated how retinal size information affects judgments of disparity-defined depth (i.e. the depth specified by relative disparity and information about the absolute distance from vergence). The premise of this work was as follows. Under monocular viewing, the retinal size of an unfamiliar target is ambiguous with respect to its distance: for any given retinal projection there are infinite combinations of physical sizes and distances that can give rise to the projected image. As a result, the visual system should operate under the assumption that the cue is not a good determinant of distance, so that when a reliable estimate of depth is available from disparity, retinal size should not affect judgments of disparity-defined distance - especially at short distances, where disparity information is thought to be highly effective (Howard & Rogers, 2002).

This idea was first investigated by measuring psychometric functions for disparity-defined depth in a number of experiments. In the first experiment observers saw either two targets of equal retinal size or two targets of different retinal size. When the two targets had equal

retinal sizes, observers accurately judged the depth between them (i.e. there was no shift in the PSE). Surprisingly, when retinal sizes differed between targets, observers were systematic biased, such that a target with a larger retinal size was reported to be closer than a target with a smaller retinal size at the same disparity-defined distance. Two follow-up experiments showed that (1) this bias increased as the ratio between the retinal sizes increased; and (2) the absolute distance to both targets was increased.

We then showed that a Bayesian (probabilistic) cue combination model can account for our results reasonably well. To construct this model, we first considered how the visual system may judge depth from the combination of disparity and vergence. To implement this model, we first constructed a probability density function for disparity as a function of the absolute and relative distance. This probability function shows that a given measurement of disparity cannot specify a single value for absolute or relative distance. However, we could also specify a probability distribution for vergence, which can only specify the absolute distance but not relative distance. When Bayes' Rule is used to multiply these two distributions, the posterior distribution specifies a single likely value of absolute and relative distance, which can then be used to calculate the metric relative depth between two targets. In this manner, the model was shown to accurately predict the actual disparity-defined depth that is specified for a given range of disparities at a given viewing distance.

This model was then extended to account for biased reports of depth when retinal size information was included. In particular, we proposed that the visual system uses an assumption that alters the interpretation of the retinal size information, namely that the differences in the retinal projections of the two targets were produced by two objects of equal size but at different distances. Under this assumption, the ratio between the retinal sizes is inversely proportional to the relative distance ratio, even though - in reality - it is not. This assump-

tion was then used to construct a probability distribution for relative depth from the retinal size information. Using Bayes' Rule, we then multiplied the distributions for relative disparity, vergence and retinal size. The model accurately predicts that the addition of the retinal size information caused a slight shift in the posterior distribution with respect to the true disparity-defined depth and it is this shift that causes perceptual bias.

To explain the results from Experiments 2 (retinal size ratio vs. absolute size) we can then shift the mean of the retinal size distribution (so as to signal a different measured ratio between the targets' retinal sizes). Consistent with the psychophysical findings, the model predicts an increase in the perceptual bias as the retinal size ratio increases. Likewise, we can explain the results from Experiment 3 (different absolute distances) by shifting the mean of the vergence distribution (so as to signal different absolute distances). Again, consistent with our psychophysical results, an increase in the absolute distance resulted in a higher predicted perceptual bias.

The model as presented as part of this thesis predicts the qualitative properties of our results (i.e. the trends in our results) reasonably well. Quantitatively, however, the model provides a poor fit to our psychophysical data (i.e. the predicted perceptual bias does not match the measured perceptual bias). This is most likely caused by the use of different observers in each experiment. In the first experiment of the chapter, we present the results of nine observers. From these results (Figure 5.4) it is clear that there is large between-observer variability. This complicates our approach of fitting our model to each experiment using the same parameters, as different observers are likely to vary in the degree to which each cue contributes to the final estimate of depth. In addition, the standard deviations we chose to fit the model to our psychophysical results were not directly motivated by the psychophysical results and as such are free parameters; rather, we chose standard deviations that seemed reasonable,

given the nature and adjusted these to fit the results for Experiment 1. Future work will seek to reduce the number of free parameters and improve the quantitative model fits.

7.1.4 Chapter 6

In this chapter, we extended the findings from Chapter 5 to judgments of motion in depth. Specifically, we investigated the role of physical object size in judging the speed of three-dimensional motion (i.e. motion towards the observer). Differences in the physical size of two targets approaching at the same speed on the same trajectory cause differences between the looming rates and the retinal sizes of the targets. However, just as retinal size is an ambiguous cue to distance, monocular looming information is ambiguous with respect to the 3D speed of an approaching target. As a result, we expected that differences between the looming rates of two approaching targets should not affect judgments of 3D speed; changes in disparity can provide an estimate of the 3D speed independently from differences in physical object size.

To investigate this, we sequentially presented two approaching targets (outline circles) over similar motion trajectories. In one condition we presented two targets of equal physical size, in another we presented two targets of differing size. As a result, looming and retinal size information differed, but the disparity information was equal for both conditions. Our results showed that target size systematically affected judgments of 3D speed, so that a larger target was reported to move faster than a smaller target that moved at the same speed over the same trajectory. Conversely, when we presented two targets of equal size, judgments were unbiased.

To account for these findings, we attempted to incorporate the findings from Chapter 5. In particular, we argued that a target with a smaller retinal size (i.e. here a physically smaller

target) is likely to be seen as farther - consistent with the results from Chapter 5 - leading to a biased estimate of the distance to the target. If this distance is then used to scale the changing disparity signal, this may have accounted for our results. However, from binocular geometry it follows that when an object is seen as farther its 3D speed should be perceived as faster, meaning that a smaller target (which is seen as farther) would be seen as faster. Our results show the opposite: a smaller target is reported to move slower thus this cannot account for our results. Rather, we suggest that perceptual bias is due to an interaction of biased viewing distances from differences in retinal size and differences in looming rate. Future work will aim to investigate how observers may recover the 3D speed from looming signals.

7.2 Which information is used?

Throughout this thesis I have asked which sources of information the visual system uses to estimate depth and motion-in-depth. Here I discuss how the results presented in this thesis contribute to an understanding of the visual system's use of perceptual information.

In Chapter 3, we showed that, despite high correlations of time-to-contact (TTC) with other perceptual variables (e.g. approach speed or displacement), observers seem to judge TTC rather than its covariates when asked to judge TTC. This suggests that the visual system uses the perceptual information to recover a temporal estimate. Furthermore, it has long been shown that looming provides a strong cue to motion in depth and it is generally agreed that observers can make a perceptual judgment exclusively on the basis of looming (e.g. Lee, 1976; Regan & Hamstra, 1993; Lopez-Moliner, Field, & Wann, 2007) under constrained circumstances (Wann, 1996). However, when monocular and binocular cues are available, the visual system uses both (Gray & Regan, 1998; Rushton & Wann, 1999), a finding that is also reflected in our results. In particular, we find that the looming rate did not provide a good

account of our results - especially for Experiment 2. Rather, judgments were likely to be based on a combination of binocular and monocular information. This finding provides indications that the visual system uses all available perceptual information to estimate 3D motion.

More evidence that the visual system uses all available information was found in Chapter 4, in which we investigated the contribution of extra-retinal signals to changes in vergence to judgments of 3D speed (i.e. the speed of motion towards and away from the observer). Traditional views maintain that extra-retinal cues do not contribute to perceptual judgments of 3D motion. Recent work (Welchman et al., 2009) suggested that this claim is false: they showed that judgments of sign of motion (i.e. approaching or receding) were best explained on the basis that observers combine retinal and extra-retinal signals to 3D motion.

In Chapter 4, we ask whether extra-retinal cues support judgments of 3D speed as well as motion sign: extra-retinal signals were shown to systematically affect judgments, so that when the eyes were moving, an approaching target was reported as faster during convergence and slower during divergence. This provides further evidence that - contrary to widely held beliefs - extra-retinal signals do contribute to judgments of 3D motion. Although its contribution is small with respect to the contribution of changing disparity information, we show that its contribution cannot be dismissed. As is the case with TTC judgments, we show that the visual system uses all available information, even though the relative contribution of one cue may be minor.

In addition to this main finding, we also found that when the eyes were maximally converged (near distance) or diverged (far distance) there judgments of 3D speed were not significantly different. That is to say, 3D speed judgments were uncorrected for the viewing distance, whereas binocular geometry dictates that at a far distance the target should be perceived as approaching faster. This was not the case and suggests that the visual system represents 3D

speed in terms of angular units, rather than recovering the real-world 3D speed.

In Chapter 5, we found that judgments of depth between two targets were systematically affected by differences in retinal size. This finding is interesting, as retinal size information is inherently ambiguous with respect to depth and thus should not inform judgments of depth. However, here we show that this result is due to the visual system reinterpreting the retinal size information by assuming that the presented targets were of equal size - even though in reality they were not. As a result, the visual system used the relative size between the objects as a cue. In Chapter 6, we extend these findings to 3D motion: a target's physical size systematically affected judgments of 3D speed. Again, this finding is interesting, as looming rate (the rate of change in the retinal size of a target) should not inform judgments of 3D speed without knowing the physical target size or the absolute distance to the target. In these situations, where two cues signal significantly different depths, it may be more sensible for the visual system to exclude the retinal size or looming cue from the perceptual decision making process (more about this in the next section) and base their judgments on other cues (e.g. relative disparity or changing disparity cues). Here, we show this is not the case and the visual system seems to use the retinal size and looming cues, even though they induce bias.

Summarising, the results from all experimental chapters suggest that the visual system incorporates all available perceptual information into judgments of both static depth and motion-in-depth.

7.3 How are depth cues combined?

The current dominating theoretical cue combination is one that is based on Bayesian statistics. In particular, it is thought that the visual system combines sensory information on the basis of an optimality principle, which combines information so that the variance (i.e. the uncertainty)

of the cue-combined estimate is minimized (Landy et al., 1995). Bayesian approaches to cue combination have successfully accounted for many aspects of depth perception, including surface slant (Hillis et al., 2002, 2004; Knill & Saunders, 2003), object shape (Ernst & Banks, 2002; Johnston, 1991; Johnston et al., 1993) and distance perception (e.g. Held et al., 2010; Svarverud et al., 2010). In this thesis we applied Bayesian cue combination approaches to account for the results from Chapters 4 and 5. These are reviewed below.

In Chapter 4, we showed that the visual system optimally combines retinal and extra-retinal signals using a weighted average approach in which each cue's contribution is dictated by their relative reliabilities. The model we present in this chapter accounts extremely well for our results - considering that there were no free parameters. In Chapter 5 we provided evidence that the visual system uses a Bayesian framework to combine information from relative disparity and retinal size. This approach is an elegant solution to cue combination, as it allows for combination of relative and absolute distance information and incorporates the uncertainty that is inherent to each cue. Interestingly, however, the equal-size assumption that the visual system appears to make when interpreting the retinal size information introduces a large discrepancy between the depth signalled by relative disparity and that signalled by retinal size; it is this discrepancy that leads to perceptual bias. It has previously been suggested that the visual system may operate in a robust manner and veto (i.e. discard) the biased cue (Landy et al., 1995; Blake et al., 1993; Banks & Backus, 1998). Yet, our results seem to suggest that the visual system does not veto retinal size information: the biased information is taken into account and causes perceptual bias. This shows that the optimisation of cue combination with respect to *just* the variance may come at the cost of perceptual bias.

Summarising, the results of these chapters seem to suggest that - consistent with contemporary views - the visual system optimally combines different sources of sensory information,

even though this approach can sometimes lead to perceptual bias.

7.4 Conclusions

In this thesis I have addressed two related questions. First, which sources of information does the visual system use? Here, I showed that when we ask people to judge time-to-contact, they base their judgments on the actual time-to-contact, rather than on one or more of its covariates. Our results indicate that observers do not rely only on monocular information to looming, but rather judge TTC on the basis of both binocular and monocular information. In addition, we showed that extra-retinal signals are incorporated in judgments of 3D speed, even though observers cannot appreciate 3D motion from extra-retinal signals in isolation. Finally, we showed that although monocular retinal size information is inherently ambiguous, observers used this information to judge the depth between two objects of different retinal size; our results indicated that, to interpret the retinal size information, they make assumptions about the physical size of objects. We then extended this finding to judgments of 3D speed: observers appear to use differences between the looming signals (changes in retinal size) of two approaching targets, even though monocular looming signals cannot uniquely signal 3D speed without additional, but false, assumptions. These results suggest that the visual system uses all available information and does not discard information that is potentially uninformative or inherently ambiguous.

Second, when multiple cues to depth are available, how does the visual system then combine these into a unified percept of depth? The current dominant cue combination theory is based on Bayesian statistics, an approach that combines information based on an optimality principle that optimises the variance of the final estimate. In two of our chapters, we present computational models that show that the visual system uses this cue combination strategy in

making decisions about both 3D motion and static depth. For example, we found that the visual system optimally combines retinal and extra-retinal signals to 3D motion. In addition, we found good evidence that information to relative retinal size and relative disparity is optimally combined, even though this leads to systematic perceptual bias. This indicates that optimal combination may not always be a sensible strategy to take, especially when the available information requires additional assumptions to inform judgments; assumptions that may not always be true.

A

Assessing stereoacuity in naive observers using random dot stereograms

A.1 introduction

The angles from rays of light emanating from a near object are larger than those produced by rays from an object that is farther. The relative disparity between these angles provides the principal stimulus for the discrimination of depth between the objects. Sensitivity to this angular disparity, is known to be exquisite: experienced adult observers can, under optimal conditions, discriminate only a few seconds of arc (e.g. Kumar & Glaser, 1994; Westheimer, 1979). These stereo discrimination thresholds are often taken as an index for stereoacuity: how well depth is seen from stereo. However, the magnitude of stereo thresholds is known to depend on the testing protocol and the properties of the stimulus. For example, Westheimer (1979) showed that there is a large difference between stereo thresholds for sequentially and simultaneously presented targets: stereo thresholds for sequentially presented stimuli are typically tenfold those found for simultaneously presented targets. This may be due to random variations in measurements in the positions of the eyes before and after an eye movement is made. Conversely, when objects are presented simultaneously, the direct comparison of the stimuli essentially reduces the task of judging disparity to a vernier task (Foley, 1976).

Other factors that influence the magnitude of stereo thresholds are: (1) the duration between stimulus presentation (Foley, 1976; Westheimer, 1979); (2) spatial acuity (Bradshaw & Rogers, 1996); and (3) contrast (Harwerth, Fredenburg, & Smith, 2003). Finally and foremost, stereo thresholds are believed to depend on normal visual acuity of the two eyes and correct ocular alignment. Stereoacuity test are often administered to diagnose binocular abnormalities, such as anisometria (the condition where the two lenses have unequal refrative power), amblyopia (poor or lack of vision in an eye that is otherwise physically norma) and strabismus (a condition in which the the two eyes are not aligned properly).

For this thesis, I conducted experiments whose results often critically depended on the observers' ability to see depth from stereo (i.e. relative disparity). However, not everyone is able to recover depth from stereo reliably. Richards (1970, 1971) reported that about 30% of people have difficulty detecting either crossed or uncrossed disparity in stationary line stereograms. About 3% of observers is completely unable to recover depth from stereo and is classed as 'stereo blind'. Due to the idiosyncratic nature of stereo vision, it is useful to assess whether observers can recover depth from stereo but also how well they can see stereo (i.e. what the smallest disparity is that they can discriminate).

Many stereoacuity tests are available commercially; the TNO test (Ootech, AG Veenendaal, the Netherlands), the Randot test and the Titmus (Stereo Optical Co.,Inc, Chicago, IL, USA) are amongst the three most frequently used tests - both in clinical settings and in vision research. However, limitations to these tests exist. For example, the TNO test presents stereograms through red-green anaglyphs, but using anaglyphs can induce rivalry between the left- and right eyes' images, cross-talk and aniseikonia (differences in the perceived image size) - each of which can reduce stereo thresholds. Finally, most tests of stereoacuity use the (ascending) method of limits, which is similar to how visual acuity is measured. To illustrate,

a person taking the Snellen acuity test covers one eye and reads aloud rows of characters. The row containing the smallest characters that can be read out accurately is taken as the index of visual acuity. Similarly, when stereoacuity is measured observers are sequentially presented with relative disparities that decrease in size; the smallest disparity that can be identified accurately is then taken as the stereo threshold. However, unlike visual acuity the disparity function is signed: objects can be in front of (crossed disparity) or behind (crossed disparity) other objects or fixation. Any test that exclusively tests crossed or uncrossed disparity does not cover the full range of performance.

Given the limitations of most existing tests, I developed a simple near/far disparity discrimination test for initial screening of naive subjects. In this test I used random dot stereograms, which were popularised as a tool for research by Julesz (1960) and subsequently used in many human and primate stereopsis experiments (e.g. DeAngelis, 2000; Harwerth & Boltz, 1979; Schor, 1991). This stimulus operates in the cyclopean domain; i.e. to see depth in a random dot stereogram requires the cooperation of the two eyes. As a result, when a random dot stereogram is seen monocularly (i.e. with one eye) the observer should not be able to see depth but rather a flat array of randomly positioned dots. This makes this stimulus an ideal stimulus to investigate relative disparity in isolation.

Prompted by a post to VisionList by Prof. Julie Harris ("Observer rejection query"; January 13, 2010) I reviewed the results of the screening of 84 naive participants. Here, I present these results as well as characteristic response patterns and I discuss the test's usability to describe binocular function.

A.2 Methods

A.2.1 Apparatus

Stimuli were presented stereoscopically on a 2-monitor haploscope (see General Methods) at a viewing distance of 50 cm, with inter-ocular spacing and vergence angles appropriately configured for each observer. The resolution of the screen was 1600 by 1200 pixels (each pixel subtending about 1.75 by 1.75 arcmin). Stimuli were programmed in C#, using OpenGL image libraries.

A.2.2 Stimulus

The stimulus (see Figure A.1 for a cartoon) consisted of a random dot stereogram (the pedestal stimulus; 18 deg by 13.5 deg) in which a central area (the test stimulus; 7.5 deg by 7.5 deg) was displaced in depth (range ±9 arcmin spaced in unequal steps around 0 arcmin). The

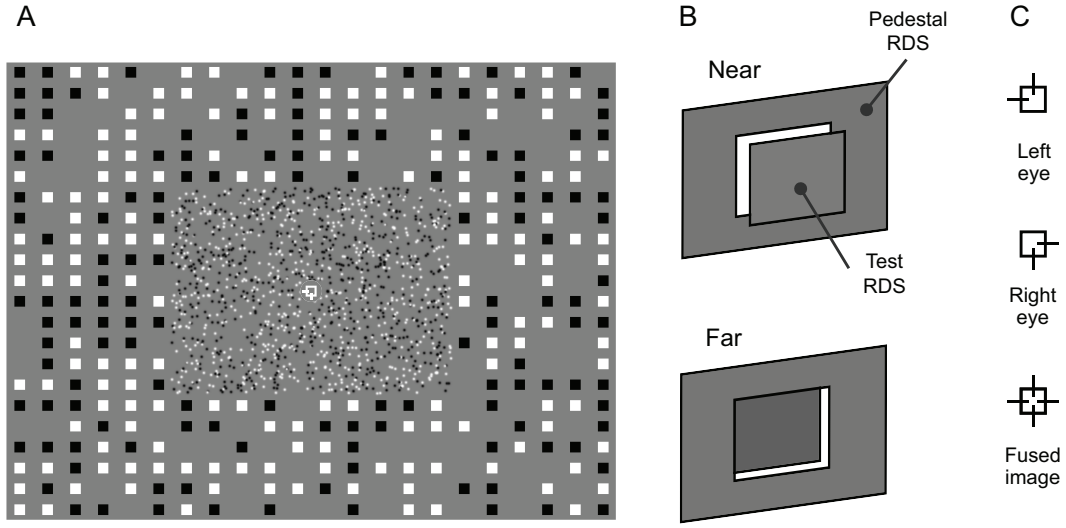

Figure A.1: Illustration of the stimulus. **(A)** One monocular half image of a stereogram, showing a background of squares, the reference patch and the test patch (not outlined, but contained within the reference patch). The nonius lines are shown in the middle, with a grey circular background. **(B)** A cartoon illustrating the instruction for observers (background not shown). The outer square is the pedestal RDS; the inner square is the test RDS, which was varied in depth with respect to the pedestal which was always in the plane of the screen (i.e. $\delta = 0$ arcmin). The upper panel shows a "near "stimulus; the lower panel shows a far stimulus. **(C)** The nonius line arrangement. When vergence is appropriate, the left and right eyes' images (top and middle panels, respectively) fuse into a box with a cross thought it (bottom panel).

random dot stereogram was made up of Gaussian dots (radius about 5 arcmin) on a gray background and was surrounded by a pattern of squares (side length about 21 arcmin). To promote correct fusion of the half-images about half of the dots in the stereogram and squares in the background was white, the rest were black. Dot density was 20 dots per square degree. A set of nonius lines was presented in the centre of the test RDS to promote correct vergence.

Displacement in depth of the test stimulus was simulated by varying its relative disparity with respect to the pedestal stimulus. This relative disparity was generated by laterally shifting the test stimulus in opposite directions in the two eyes relative to the pedestal stimulus. To illustrate, the method of creating the stimulus is analogous to sliding two layers of random dots over each other, the top layer containing a smaller area than the bottom layer. In this manner, columns of random dots from the bottom layer will automatically result in appropriate monocular zones in each eye and there is no need to move any dots into empty monocular zones as the traditional methods have done. To calculate the amount by which the test patch should be shifted in each eye I first calculated the offset in depth (Δ) generated by a value of disparity (δ):

$$\Delta = \frac{\delta z^2}{i - \delta z} \tag{A.1}$$

where i is the inter-ocular separation and z is the distance to the screen (see Appendix A). I then used the resulting depth offset to calculate the x-offset (x) for each eye:

$$x = \frac{0.5i\Delta}{z + \Delta}, \tag{A.2}$$

where the sign of x was inverted for the left and right eyes: when a dot moved to the left in the left eye (i.e. uncrossed disparity) the dot was moved to the right in the right eye.

A.2.3 Procedure

Observers sat in a dark laboratory and briefly viewed (300 ms) a random dot stimulus on each trial. This short presentation duration is typically enough to prevent vergence eye movements, thereby ensuring that judgments were based on relative disparity; Westheimer & Mitchell, 1969. Observers were instructed to fixate the nonius lines such that they formed a box with a cross through it (Figure A.1C). The pedestal was always located in the plane of the screen; The disparity of the test patch relative to the pedestal was controlled using the method of constant stimuli. On each trial, observers indicated whether they saw the test stimulus in front of or behind the pedestal, after which there was a pause of 500 ms before the next trial was presented. Each level of disparity was presented 20 times (9 levels of disparity; 180 trials in total). A single test was typically completed within 3 minutes.

A.2.4 Observers

The data reported here were collected from prospective observers in experiments conducted in the binocular vision laboratory at the University of Birmingham. Prior to participation they were first asked to fill out a form to indicate their visual acuity and visual defects. Applicants with a correction of ±5 or known visual disorders were excluded from screening.

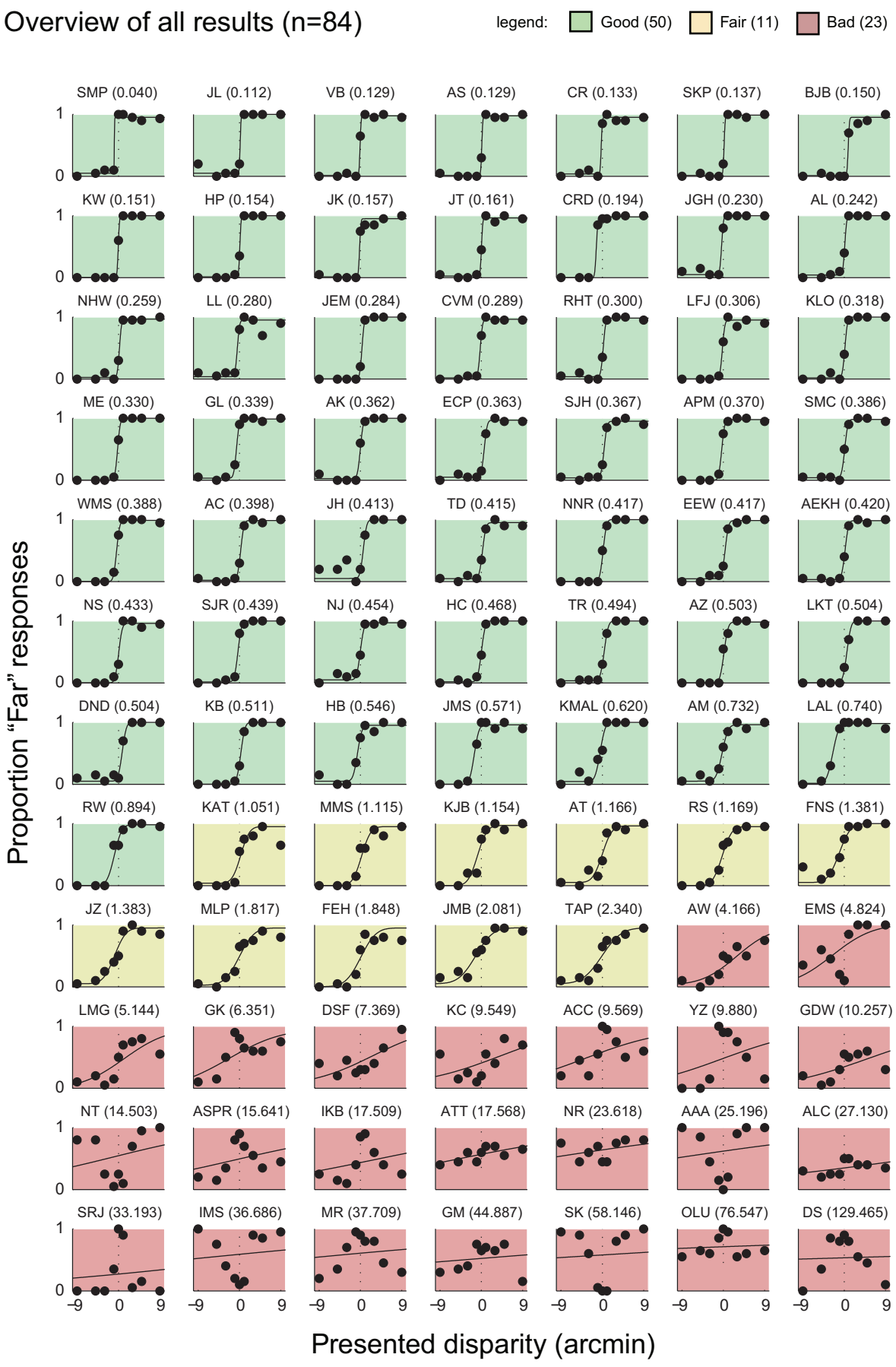
Overview of all results (n=84)
legend: Good (50) Fair (11) Bad (23)
SMP (0.040) JL (0.112) VB (0.129) AS (0.129) CR (0.133) SKP (0.137) BJB (0.150)
KW (0.151) HP (0.154) JK (0.157) JT (0.161) CRD (0.194) JGH (0.230) AL (0.242)
NHW (0.259) LL (0.280) JEM (0.284) CVM (0.289) RHT (0.300) LFJ (0.306) KLO (0.318)
ME (0.330) GL (0.339) AK (0.362) ECP (0.363) SJH (0.367) APM (0.370) SMC (0.386)
WMS (0.388) AC (0.398) JH (0.413) TD (0.415) NNR (0.417) EEW (0.417) AEKH (0.420)
NS (0.433) SJR (0.439) NJ (0.454) HC (0.468) TR (0.494) AZ (0.503) LKT (0.504)
DND (0.504) KB (0.511) HB (0.546) JMS (0.571) KMAL (0.620) AM (0.732) LAL (0.740)
RW (0.894) KAT (1.051) MMS (1.115) KJB (1.154) AT (1.166) RS (1.169) FNS (1.381)
JZ (1.383) MLP (1.817) FEH (1.848) JMB (2.081) TAP (2.340) AW (4.166) EMS (4.824)
LMG (5.144) GK (6.351) DSF (7.369) KC (9.549) ACC (9.569) YZ (9.880) GDW (10.257)
NT (14.503) ASPR (15.641) IKB (17.509) ATT (17.568) NR (23.618) AAA (25.196) ALC (27.130)
SRJ (33.193) IMS (36.686) MR (37.709) GM (44.887) SK (58.146) OLU (76.547) DS (129.465)
Proportion "Far" responses
0
1
−9 0 9
Presented disparity (arcmin)

Figure A.2: (previous page) Stereo test results (thresholds) for 84 observers. Results are ordered by their respective thresholds and divided into three categories: Good (green panels; threshold < 1 arcmin), Fair (yellow panels; threshold between 1 and 2.5 arcmin) and Bad (red panels; threshold > 2.5 arcmin).

A.3 Results

To quantify the results of each test, I first constructed psychometric functions by plotting the proportion of "behind" answers as a function of the presented disparity. I then fitted a cumulative Gaussian and used its standard deviation as a measure of the stereo threshold. Likewise, I took the shift in the PSE (the offset of the fitted Gaussian on the x-axis) as a measure of systematic bias. An overview of the results of 84 observers are shown in Figure A.2, in which each panel shows a single result from a unique observer (i.e. there are no double sessions in the overview). Of 84 observers, 50 showed thresholds under 1 arcmin (Figure A.2, green panels). A further 11 observers showed thresholds between 1 and 2.5 arcmin (Figure A.2, yellow panels) and 23 showed thresholds over 2.5 arcmin (Figure A.2, red panels). The mean threshold was 0.5±1.23 arcmin; the mean bias was -0.036±1.1 arcmin (Figure A.3).

A.4 Discussion

There are large individual differences in stereoacuity between observers. It is known that human observers, under optimal conditions, can discriminate between a few arcs of disparity. However, the measurements that lead to these low values are often produced by experienced observers. Here, I present data from 84 naive psychophysical observers. The average discrimination threshold for these observers was about 30 seconds of arc. Although higher than the low values commonly reported from experienced observers, this is still a low value considering that the data shown here was produced by naive observers, many of whom had never participated in psychophysical experiments.

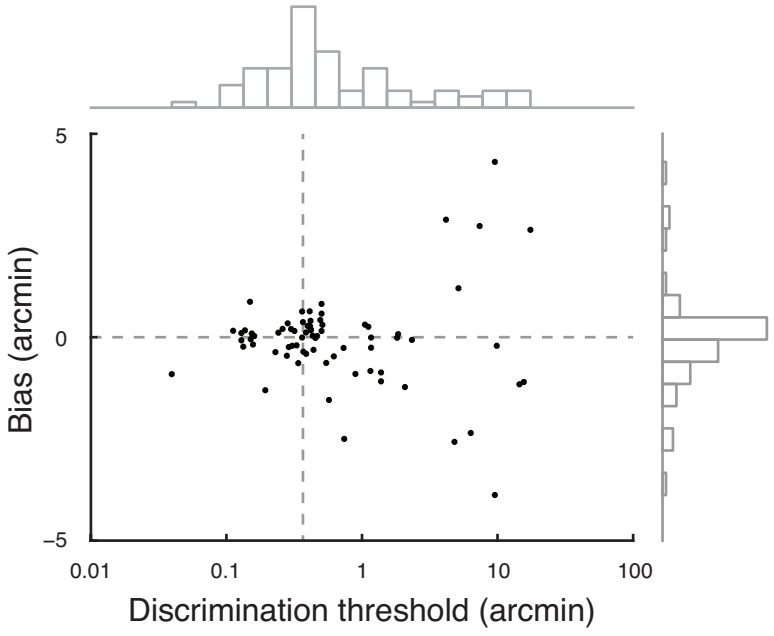

Figure A.3: Summary of stereo test results. The horizontal histogram shows the distribution of thresholds; the vertical histogram shows the distribution of biases. Gray dashed lines show the mean threshold (vertical) and the mean bias (horizontal).

A.4.1 Characteristic response patterns

When observers were sufficiently sensitive to the disparity in the RDS stimuli, I would expect a psychometric function which remained 0 at the negative values of disparity (i.e. they did never perceived the stimulus as far) and for the positive values saturated to a proportion of 1 (i.e. they would always perceive this the stimulus as far). For about 62 of 84 observers (about 75%) we found reasonable psychometric functions. However, there could be other (consistent) response patterns which bore no resemblance to a sigmoid function. First, performance could essentially be at chance level (e.g. observers ATT, ALC and NR in Figure A.2); this might indicate a degree of stereo-blindness. Another pattern which was fairly consistent was one in which the psychometric function for crossed or uncrossed disparity was normal, but in the opposite direction reduced to chance level (e.g. observers AW and DFS in Figure A.2); this may indicate that observers can see either crossed or uncrossed normally, but are impaired in seeing the opposite disparity (uncrossed or crossed, respectively). This may indicate a visual anomaly, such as strabismus or amblyopia (cf. Richards, 1971) (Although anecdotal evidence, both observers AW and DFS reported - after taking part in the test - that as a child a patch had

placed over their unimpaired eye). This may suggest that this pattern of results is consistent with amblyopia, but more clinical testing would be required to verify this claim.) Finally, I consistently found a pattern in which observers (n=11) appeared to respond on the basis of some disparity being present, regardless of its sign. These response patterns were consistent with a 'V' or 'inverted V' pattern in the psychometric function. We are unsure as to what may have led to this pattern. However, repeated testing with two of these subjects (data not shown) indicated that this results may be caused by the short presentation time of the stimulus; when we increased the presentation duration, these observers were able to recover the depth in the RDS reliably.

A.4.2 Perceptual learning in repeated exposure to RDS

We show that measuring stereoacuity by sampling a full psychometric using a briefly flashed random dot stereogram can be used adequately to assess potential observers' performance on tasks involving relative disparity. In addition, it is possible that this test may also be used to identify some visual disorders, which prevent normal stereopsis. The test is fast and provides an indication of performance on a large range of disparities. However, as pointed out by several responders to the query from Prof. Julie Harris, there are a number of disadvantages to using random dot stereograms. For example, Prof. Marty Banks replied the following:

> "We find that a lot of inexperienced observers can't initially see depth in random-element stereograms. They see them as flat. Not sure what percentage, but I wouldn't be surprised if it's 25%. With practice, the great majority learn to see depth just fine."

Indeed, it has been shown that there is a large degree of perceptual learning in these stimuli; repeated observation of RDS stimuli is typically associated with a reduction in the time to perceive the figure it contains and the depth it specifies (Julesz, 1971; Ramachandran & Brad-

dick, 1973; J. Frisby & Clatworthy, 1975; O'Toole & Kersten, 1992) and a substantial reduction in stereothresholds (Gantz, Patel, Chung, & Harwerth, 2007).

Although testing for perceptual learning was beyond the scope of this test, there are indications for increases in performance with repeated exposure to the stimuli. Specifically, due to potential learning of the stimulus, observers who initially did not perceive depth (e.g. the results did not bore any resemblance to a psychometric function) were tested again. Besides the 84 observers reported here, I have data for 13 observers who were tested more than once in succession. Most of these observers (n=8) showed the 'V' or inverted 'V' pattern discussed in the previous section. Of these observers, six improved so that their stereothreshold was below 1 arcmin of disparity. This may *not* provide indications of visual disorders, but it is more likely that learning took place, resulting in a lower presentation duration required to see depth in the RDS. For this reason it is sensible to retest observers whose responses do not show a psychometric function, especially when they show the characteristic 'V' pattern. Their performance may improve substantially in a second or third session.

A.4.3 Relation to other measures of stereoacuity

An obvious weakness pertaining to the test is the lack of direct comparisons between this test and others, therefore undermining the test's validity. However, it would be hard to compare the results from this test to the results of other tests. Most traditional stereotests use the method of limits and measure an absolute threshold, which - although related - is not the same measure for stereoacuity as the discrimination threshold we used as a measure in this test. Furthermore, it is well known that the reported stereoacuity depends greatly on the method used to measure it (e.g. Heron, Dholakia, Collins, & McLaughlan, 1985). In the latter, mean stereothresholds were reported to be 12.9 (Frisby Test), 28.1 (Randot), 26.6 (Titmus

Test) or 88.7 seconds of arc (TNO test). One study which investigated stereoacuity using RDS stimuli over a large range of disparity values (but used liquid-crystal shutter glasses for presentation of stereoscopic stimuli) reported stereothresholds of about 37 arcsec for observers aged 15 to 39. This is very similar to the average threshold obtained using the test reported here, supporting some validity of the test.

A.4.4 Conclusion

I conclude that the automated random dot stereotest that I developed is an adequate tool to provide initial assessment of observers' stereoacuity. In addition, there are some indications that this test can be used to identify visual disorders, such as strabismus or amblyopia; however, investigating this claim would require more thorough testing of observers who are clinically diagnosed with these disorders.

Acknowledgements

I thank Matthew Patten, Aidan Murphy, Hiroshi Ban, Dicle Dovencioglu and Christa van Mierlo for help with the collection of the data presented in this section.

B

Geometry of binocular vision

To calculate the binocular disparity (δ) between fixation (F) and point P:

$$\delta = \theta_L - \theta_R = \phi_P - \phi_F \tag{B.1}$$

$$\angle\phi_P = \frac{2j}{z} \quad \angle\phi_P = \frac{2j}{z+\Delta} \tag{B.2}$$

$$\delta = \frac{2j}{z} - \frac{2j}{z+\Delta} = \frac{2j(z+\Delta) - 2j(z)}{z(z+\Delta z)} \tag{B.3}$$

$$\delta = \frac{2j(z) + 2j(\Delta) - 2j(z)}{z(z+\Delta)} \tag{B.4}$$

$$\delta = \frac{2jz}{z^2 + z\Delta} = \frac{i\Delta}{z^2 + z\Delta} \tag{B.5}$$

Because normally $\Delta \ll z$, we can write:

$$\delta \approx \frac{i\Delta}{z^2} \tag{B.6}$$

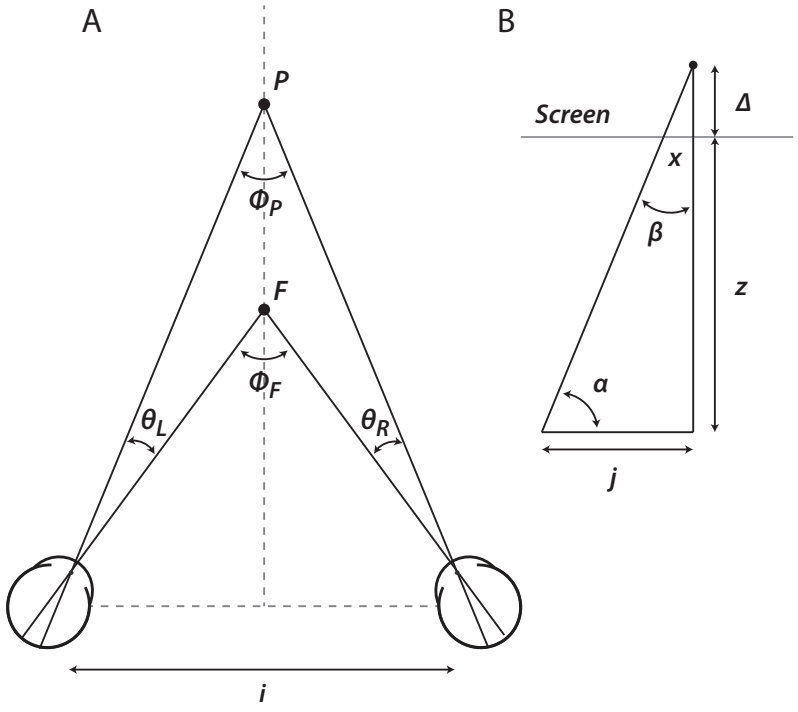

Figure B.1: Illustration of the binocular viewing geometry. (**A**) The relative binocular disparity ($\theta_L - \theta_R$) between points P and F can be calculated by taking the difference between the vergence angles. (**B**) calculating depth (Δ) with respect to the screen from the lateral offset (x).

To calculate real-world depth (Δ) from x-offset (x) on the screen:

$$\frac{\Delta}{x} = \frac{z + \Delta}{j} \tag{B.7}$$

$$\Delta j = xz + x\Delta \tag{B.8}$$

$$\Delta = \frac{xz}{j - x} \tag{B.9}$$

To calculate the x-offset (x) in each eye on the screen from disparity:

First we derive the depth (Δ) from disparity (δ, in radians):

$$\delta = \frac{i\Delta}{z^2} \Leftrightarrow \Delta = \frac{\delta z^2}{i} \tag{B.10}$$

Using Equation A.9, we can write:

$$\frac{xz}{0.5i - x} = \frac{\delta z^2}{i} \tag{B.11}$$

$$xzi = 0.5i\delta z^2 - x\delta z^2 \tag{B.12}$$

$$x(zi + \delta z^2) = 0.5i\delta z^2 \tag{B.13}$$

$$x = \frac{i\delta z^2}{2(zi + \delta z^2)} \tag{B.14}$$

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.*, *14*, 257-262.

Ames, A., & Gliddon, G. H. (1928). Ocular measurements. *Transactions of the Section of Ophthalmology, AMA*, 102-175.

Backus, B. T., Banks, M. S., van Ee, R., & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Res*, *39*, 1143-1170.

Backus, B. T., & Matza-Brown, D. (2003). The contribution of vergence change to the measurement of relative disparity. *J Vis*, *3*, 737-750.

Banks, M. S., & Backus, B. T. (1998). Extra-retinal and perspective cues cause the small range of the induced effect. *Vision Res*, *38*, 187-194.

Beverley, K. I., & Regan, D. (1973). Evidence for the existence of neural mechanisms selectively sensitive to the direction of movement in space. *Journal of Physiology*, *235*, 17-29.

Bishop, P. O. (1989). Vertical disparity, egocentric distance and stereoscopic depth constancy: a new interpretation. *Proc. R. Soc. Lond., B, Biol. Sci.*, *237*, 445-469.

Blake, A., Bulthoff, H. H., & Sheinberg, D. (1993). Shape from texture: ideal observers and human psychophysics. *Vision Res.*, *33*, 1723-1737.

Blakemore, C. (1970). The range and scope of binocular depth discrimination in man. *J. Physiol. (Lond.)*, *211*, 599-622.

Bootsma, R. J., & Wieringen, P. C. W. van. (1990). Timing an attacking forehand drive in table tennis. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 21-29.

Bradshaw, M. F., Glennerster, A., & Rogers, B. J. (1996). The effect of display size on disparity scaling from differential perspective and vergence cues. *Vision Res*, *36*, 1255-1264.

Bradshaw, M. F., & Rogers, B. J. (1996). The interaction of binocular disparity and motion parallax in the computation of depth. *Vision Res*, *36*, 3457-3468.

Brenner, E., Smeets, J. B., & Landy, M. S. (2001). How vertical disparities assist judgements of distance. *Vision Res*, *41*, 3455-3465.

Brenner, E., & van Damme, W. J. (1998). Judging distance from ocular convergence. *Vision Res*, *38*, 493-498.

Brenner, E., & van den Berg, A. V. (1994). Judging object velocity during smooth pursuit eye movements. *Exp Brain Res*, *99*, 316-324.

Brenner, E., Van Den Berg, A. V., & Van Damme, W. J. (1996). Perceived motion in depth. *Vision Res*, *36*, 699-706.

Bruno, N., & Cutting, J. E. (1988). Minimodularity and the perception of layout. *J Exp Psychol Gen*, *117*, 161-170.

Bulthoff, H. H., & Mallot, H. A. (1988). Integration of depth modules: stereo and shading. *J Opt Soc Am A*, *5*, 1749-1758.

Burge, J., Fowlkes, C. C., & Banks, M. S. (2010). Natural-scene statistics predict how the figure-ground cue of convexity affects human depth perception. *J Neurosci*, *30*, 7269-7280.

Caljouw, S. R., Kamp, J. van der, & Savelsbergh, G. J. (2004). Catching optical information for the regulation of timing. *Exp Brain Res*, *155*, 427-438.

Cavallo, V., & Laurent, M. (1988). Visual information and skill level in time-to-collision estimation. *Perception*, *17*, 623-632.

Champion, R. A., & Freeman, T. C. A. (2010). Discrimination contours for the perception of head-centered velocity. *J Vis*, *10*, 14.

Clark, A. L., J. J. & Yuille. (1990). *Data fusion for sensory information processing systems*. Boston: Kluwer Academic Publishers.

Collett, T. S., Schwarz, U., & Sobel, E. C. (1991). The interaction of oculomotor cues and stimulus size in stereoscopic depth constancy. *Perception*, *20*, 733-754.

Coull, J. T., Vidal, F., Goulon, C., Nazarian, B., & Craig, C. (2008). Using time-to-contact information to assess potential collision modulates both visual and temporal prediction networks. *Front Hum Neurosci*, *2*, 10.

Cumming, B. G., Johnston, E. B., & Parker, A. J. (1991). Vertical disparities and perception of three-dimensional shape. *Nature*, *349*, 411-413.

Cumming, B. G., & Parker, A. J. (1994). Binocular mechanisms for detecting motion-in-depth. *Vision Res*, *34*, 483-495.

Cutting, J., & Vishton, P. (1995). Handbook of perception and cognition: Vol. 5: Perception of space and motion. In W. Epstein & S. Rogers (Eds.), (chap. Perceiving layout: The integration, relative dominance, and contextual use of different information about depth). NY: Academic Press.

DeAngelis, G. (2000). Seeing in three dimensions: the neurophysiology of stereopsis. *Trends in Cognitive sciences*, *4*, 80-90.

DeLucia, P. R. (1991). Pictorial and motion-based information for depth perception. *J Exp Psychol Hum Percept Perform*, *17*, 738-748.

DeLucia, P. R. (2005). Does binocular disparity or familiar size information override effects of relative size on judgements of time to contact? *Q J Exp Psychol A*, *58*, 865-886.

Dosher, B. A., Sperling, G., & Wurst, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Res*, *26*, 973-990.

Emmert, E. (1881). Grossenverhaltnisse der nachbilder. *Klinische Monatsblatter fur Augenheilkunde und f□r augenarztliche Fortbildung*, *19*, 443-450.

Enright, J. T. (1991). Exploring the third dimension with eye movements: better than stereopsis. *Vision Res*, *31*, 1549-1562.

Enright, J. T. (1996). Sequential stereopsis: a simple demonstration. *Vision Res*, *36*, 307-312.

Erkelens, C. J., & Collewijn, H. (1985a). Eye movements and stereopsis during dichoptic viewing of moving random-dot stereograms. *Vision Res*, *25*, 1689-1700.

Erkelens, C. J., & Collewijn, H. (1985b). Motion perception during dichoptic viewing of moving random-dot stereograms. *Vision Res*, *25*, 583-588.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429-433.

Field, D. T., & Wann, J. P. (2005). Perceiving time to collision activates the sensorimotor cortex. *Curr Biol*, *15*, 453-458.

Foley, J. M. (1976). Successive stereo and vernier position discrimination as a function of dark interval duration. *Vision Res*, *16*, 1269-1273.

Foley, J. M. (1980). Binocular distance perception. *Psychological Review*, *87*, 411-433.

Freeman, T. C. (2001). Transducer models of head-centred motion perception. *Vision Res*, *41*, 2741-2755.

Freeman, T. C., & Banks, M. S. (1998). Perceived head-centric speed is affected by both extra-retinal and retinal

errors. *Vision Res*, *38*, 941-945.

Freeman, T. C., Champion, R. A., & Warren, P. A. (2010). A bayesian model of perceived head-centered velocity during smooth pursuit eye movement. *Current Biology*, *20*, 757-762.

Freeman, T. C., & Fowler, T. A. (2000). Unequal retinal and extra-retinal motion signals produce different perceived slants of moving surfaces. *Vision Res*, *40*, 1857-1868.

Frisby, J., & Clatworthy, J. L. (1975). Learning to see complex random-dot stereograms. *Perception*, *4*, 173-178.

Frisby, J. P., Catherall, C., Porrill, J., & Buckley, D. (1997). Sequential stereopsis using high-pass spatial frequency filtered textures. *Vision Res*, *37*, 3109-3116.

Gantz, L., Patel, S., Chung, S., & Harwerth, R. (2007). Mechanisms of perceptual learning of depth discrimination in random dot stereograms. *Vision Res*, *16*, 2170-2178.

Geri, G., Gray, R., & Grutzmacher, R. (2010). Simulating time-to-contact when both target and observer are in motion. *Displays*, *31*, 59-66.

Gillam, B. (1980). Geometric Illusions. *Scientific American*, *242*, 102-111.

Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1998). Cues to viewing distance for stereoscopic depth constancy. *Perception*, *27*, 1357-1365.

Gogel, W. (1964). Size cue to visually perceived distance. *Psychological Bulletin*, *62*, 217-35.

Gogel, W. (1969). The sensing of retinal size. *Vision Res*, *9*, 1079-94.

Gogel, W., & Tietz, J. D. (1973). Absolute motion parallax and the specific distance tendency. *Perception and Psychophysics*, *13*, 284- 292.

Gogel, W., & Tietz, J. D. (1977). Eye fixation and attention as modifiers of perceived distance. *Percept Mot Skills*, *45*, 343-362.

Gogel, W., & Tietz, J. D. (1979). A comparison of oculomotor and motion parallax cues of egocentric distance. *Vision Res.*, *19*, 1161-1170.

Gonzalez, E. G., Allison, R. S., Ono, H., & Vinnikov, M. (2010). Cue conflict between disparity change and looming in the perception of motion in depth. *Vision Res*, *50*, 136-143.

Gray, R., & Regan, D. (1998). Accuracy of estimating time to collision using binocular and monocular information. *Vision Res*, *38*, 499-512.

Gray, R., & Sieffert, R. (2005). Different strategies for using motion-in-depth information in catching. *J Exp Psychol Hum Percept Perform*, *31*, 1004-1022.

Harris, J. M. (2006). The interaction of eye movements and retinal signals during the perception of 3-d motion direction. *J Vis*, *6*, 777-790.

Harris, J. M., Nefs, H., & Grafton, C. E. (2008). Binocular vision and motion-in-depth. *Spat Vis*, *21*, 531-547.

Harris, J. M., & Watamaniuk, S. N. (1995). Speed discrimination of motion-in-depth using binocular cues. *Vision Res*, *35*, 885-896.

Harwerth, R. S., & Boltz, R. L. (1979). Stereopsis in monkeys using random dot stereograms: the effect of viewing duration. *Vision Res*, *19*, 985-991.

Harwerth, R. S., Fredenburg, P. M., & Smith, E. L. (2003). Temporal integration for stereoscopic vision. *Vision Res*, *43*, 505-517.

Heinemann, E., Tulving, E., & Nachmias, J. (1959). The effect of oculomotor adjustments on apparent size. *American Journal of Psychology*, *72*, 32-45.

Held, R. T., Cooper, E. A., O'Brien, J. F., & Banks, M. S. (2010). Using blur to affect perceived distance and size. *ACM Trans Graph*, *29*.

Heron, G., Dholakia, S., Collins, D. E., & McLaughlan, H. (1985). Stereoscopic threshold in children and adults. *Am J Optom Physiol Opt*, *62*, 505-515.

Hershenson, M., & Samuels, S. M. (1999). An airplane illusion: apparent velocity determined by apparent distance. *Perception*, *28*, 433-436.

Heuer, H. (1993). Estimates of time to contact based on changing size and changing target vergence. *Perception*,

22, 549-563.

Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: mandatory fusion within, but not between, senses. *Science*, *298*, 1627-1630.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: optimal cue combination. *J Vis*, *4*, 967-992.

Hochberg, J. E., & McAlister, E. (1955). Relative size vs. familiar size in the perception of represented depth. *The American Journal of Psychology*, *68*, 294-296.

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *J Vis*, *8*, 3321-3330.

Holway, A. H., & Boring, E. G. (1941). Determinants of apparent visual size with distance variant. *The American Journal of Psychology*, *54*, 21-37.

Howard, I. P. (2008). Vergence modulation as a cue to movement in depth. *Spat Vis*, *21*, 581-592.

Howard, I. P., & Rogers, B. J. (2002). *Seeing in depth*. I. Porteous. Toronto, Canada.

Ittleson, W. (1951a). *The ames demonstrations in perception*. Princeton, NJ: Princeton University Press.

Ittleson, W. (1951b). Size as a cue to distance: Static localisation. *American Journal of Psychology*, *64*, 54-67.

Jacobs, R. A. (2002). What determines visual cue reliability? *Trends Cogn Sci*, *6*, 345-350.

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Res*, *31*, 1351-1360.

Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Res.*, *34*, 2259-2275.

Johnston, E. B., Cumming, B. G., & Parker, A. J. (1993). Integration of depth modules: stereopsis and texture. *Vision Res*, *33*, 813-826.

Julesz, B. (1960). Binocular depth perception of computer-generated patterns. *Bell System Technical Journal*, *39*.

Julesz, B. (1971). Foundations of Cyclopean Perception. *MIT Press*.

Karanka, J., Rushton, S. K., & Freeman, T. C. A. (2007). Effects of feedback on the timing of interception. *Perception*, *36*, 306-306.

Kaufman, L., & Kaufman, J. H. (2000). Explaining the moon illusion. *Proc Natl Acad Sci U S A*, *97*, 500-505.

Kilpatrick, F. P., & Ittelson, W. H. (1953). The size-distance invariance hypothesis. *Psychol Rev*, *60*, 223-231.

Kim, N. G., & Grocki, M. J. (2006). Multiple sources of information and time-to-contact judgments. *Vision Res.*, *46*, 1946-1958.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Res*, *43*, 2539-2558.

Komoda, M., & Ono, H. (1974). Oculomotor adjustments and size-distance perception. *Perc*, *15*, 353-360.

Kording, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*, 244-247.

Kumar, T., & Glaser, D. A. (1994). Some temporal aspects of stereoacuity. *Vision Res*, *34*, 913-925.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Res*, *35*, 389-412.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception*, *5*, 437-459.

Lee, D. N., & Reddish, P. (1981). Plummeting gannets: a paradigm of ecological optics. *Nature*, *293*, 293-294.

Lee, D. N., Young, D. S., Reddish, P. E., Lough, S., & Clayton, T. M. (1983). Visual timing in hitting an accelerating ball. *Q J Exp Psychol A*, *35*, 333-346.

Longuet-Higgins, H. C. (1982). The role of the vertical dimension in stereoscopic vision. *Perception*, *11*, 377-386.

Lopez-Moliner, J., Brenner, E., & Smeets, J. B. (2007). Effects of texture and shape on perceived time to passage: knowing "what" influences judging "when". *Percept Psychophys*, *69*, 887-894.

Lopez-Moliner, J., Field, D. T., & Wann, J. P. (2007). Interceptive timing: prior knowledge matters. *J Vis*, *7*, 1-8.

Lugtigheid, A., Brenner, E., & Welchman, A. (In Press). Speed judgments of three-dimensional motion incorporate extra-retinal information. *J Vis*.

Masson, G. S., Busettini, C., & Miles, F. A. (1997). Vergence eye movements in response to binocular disparity without depth perception. *Nature*, *389*, 283-286.

Mather, G., & Smith, D. R. (2000). Depth cue integration: stereopsis and image blur. *Vision Res.*, *40*, 3501-3506.

Mayhew, J. E., & Longuet-Higgins, H. C. (1982). A computational model of binocular depth perception. *Nature*, *297*, 376-378.

McKee, S. P. (1981). A local mechanism for differential velocity detection. *Vision Res*, *21*, 491-500.

McKee, S. P., & Welch, L. (1989). Is there a constancy for velocity? *Vision Res*, *29*, 553-561.

McLeod, R. W., & Ross, H. E. (1983). Optic-flow and cognitive factors in time-to-collision estimates. *Perception*, *12*, 417-423.

Michaels, C. F., Zeinstra, E. B., & Oudejans, R. R. (2001). Information and action in punching a falling ball. *Q J Exp Psychol A*, *54*, 69-93.

Mon-Williams, M., & Tresilian, J. R. (1999). Some recent studies on the extraretinal contribution to distance perception. *Perception*, *28*.

Mon-Williams, M., Tresilian, J. R., & Roberts, A. (2000). Vergence provides veridical depth perception from horizontal retinal image disparities. *Exp Brain Res*, *133*, 407-413.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, *257*, 1357-1363.

Nefs, H., & Harris, J. (2007). Vergence effects on the perception of motion-in-depth. *Exp Brain Res*, *183*, 313-322.

Nefs, H., & Harris, J. (2008). Induced motion in depth and the effects of vergence eye movements. *J Vis*, *8*, 811-816.

Ogle, K. (1950). *Researches in binocular vision*. Philadelphia, PA: Saunders.

Ono, H. (1966). Distal and proximal size under reduced and non-reduced viewing conditions. *Am J Psychol*, *79*, 234-241.

Ono, H., & Comerford, T. (1977). *"stereoscopic depth constancy" in stability and constancy in visual perception.* (E. W., Ed.). New York: John Wiley.

Oruc, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Res.*, *43*, 2451-2468.

O'Toole, A., & Kersten, D. (1992). Learning to see random-dot stereograms. *Perception*, *21*, 227-243.

Over, R. (1963). Size- and distance-estimates of a single stimulus under different viewing conditions. *The American Journal of Psychology*, *76*, pp. 452-457.

Palmer, S. E., & Brooks, J. L. (2008). Edge-region grouping in figure-ground organization and depth perception. *J Exp Psychol Hum Percept Perform*, *34*, 1353-1371.

Peper, L., Bootsma, R. J., Mestre, D. R., & Bakker, F. C. (1994). Catching balls: How to get the hand to the right place at the right time. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 591-612.

Predebon, J. (1993). The familiar-size cue to distance and stereoscopic depth perception. *Perception*, *22*, 985-995.

Ramachandran, V. S., & Braddick, O. (1973). Orientation-specific learning in stereopsis. *Perception*, *2*, 371-376.

Rashbass, C., & Westheimer, G. (1961). Independence of conjugate and disjunctive eye movements. *J. Physiol. (Lond.)*, *159*, 361-364.

Regan, D. (2002). Binocular information about time to collision and time to passage. *Vision Res*, *42*, 2479-2484.

Regan, D., & Beverley, K. (1979). Binocular and monocular stimuli for motion-in-depth: Changing-disparity and changing-size feed the same motion-in-depth stage. *Vision Res*, *19*, 1331-42.

Regan, D., Erkelens, C. J., & Collewijn, H. (1986). Necessary conditions for the perception of motion in depth. *Invest Ophthalmol Vis Sci*, *27*, 584-597.

Regan, D., & Gray, R. (2009). Binocular processing of motion: some unresolved questions. *Spat Vis*, *22*, 1-43.

Regan, D., & Hamstra, S. J. (1993). Dissociation of discrimination thresholds for time to contact and for rate of angular expansion. *Vision Res*, *33*, 447-462.

Regan, D., & Vincent, A. (1995). Visual processing of looming and time to contact throughout the visual field. *Vision Res*, *35*, 1845-1857.

Richards, W. (1970). Stereopsis and stereoblindness. *Exp Brain Res*, *10*, 380-388.

Richards, W. (1971). Anomalous stereoscopic depth perception. *J Opt Soc Am*, *61*, 410-414.

Richards, W., & Miller, J. F. (1969). Convergence as a cue to depth. *Perception and Psychophysics*, *5*, 317-320.

Rock, P. B., & Harris, M. G. (2006). Tau as a potential control variable for visually guided braking. *J Exp Psychol Hum Percept Perform*, *32*, 251-267.

Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature*, *361*, 253-255.

Rokers, B., Cormack, L. K., & Huk, A. C. (2009). Disparity- and velocity-based signals for three-dimensional motion perception in human mt+. *Nat Neurosci*, *12*, 1050-1055.

Rotman, G., Brenner, E., & Smeets, J. B. J. (2004). Mislocalization of targets flashed during smooth pursuit depends on the change in gaze direction after the flash. *J Vis*, *4*, 564-574.

Rotman, G., Brenner, E., & Smeets, J. B. J. (2005). Flashes are localised as if they were moving with the eyes. *Vision Res*, *45*, 355-364.

Rushton, S. K., & Duke, P. A. (2009). Observers cannot accurately estimate the speed of an approaching object in flight. *Vision Res*, *49*, 1919-1928.

Rushton, S. K., & Wann, J. P. (1999). Weighted combination of size and disparity: a computational model for timing a ball catch. *Nat Neurosci*, *2*, 186-190.

Schiff, W., & Detwiler, M. L. (1979). Information used in judging impending collision. *Perception*, *8*, 647-658.

Schor, C. M. (1991). Binocular sensory disorders. In D. Regan (Ed.), (Vol. 9, p. 179-218). Boston: CRC Press, Inc.

Sedgwick, H. (1986). *"space perception" in handbook of perception and human performance* (Vol. 1; K. Boff, L. Kaufman, & J. Thomas, Eds.). New York: John Wiley.

Shioiri, S., Saisho, H., & Yaguchi, H. (2000). Motion in depth based on inter-ocular velocity differences. *Vision Res*, *40*, 2565-2572.

Svarverud, E., Gilson, S. J., & Glennerster, A. (2010). Cue combination for 3d location judgements. *J Vis*, *10*, 1-13.

Taroyan, N. A., Buckley, D., Porrill, J., & Frisby, J. P. (2000). Exploring sequential stereopsis for co-planarity tasks. *Vision Res*, *40*, 3373-3390.

Todd, J. T. (1981). Visual information about moving objects. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 795-810.

Tresilian, J. R. (1995). Perceptual and cognitive processes in time-to-contact estimation: analysis of prediction-motion and relative judgment tasks. *Percept Psychophys*, *57*, 231-245.

Tresilian, J. R., & Mon-Williams, M. (2000). Getting the measure of vergence weight in nearness perception. *Exp Brain Res*, *132*, 362-368.

Turano, K. A., & Massof, R. W. (2001). Nonlinear contribution of eye velocity to motion perception. *Vision Res*, *41*, 385-395.

Wann, J. P. (1996). Anticipating arrival: is the tau margin a specious theory? *J Exp Psychol Hum Percept Perform*, *22*, 1031-1048.

Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *J Vis*, *5*, 834-862.

Welchman, A. E., Harris, J. M., & Brenner, E. (2009). Extra-retinal signals support the estimation of 3d motion. *Vision Res*, *49*, 782-789.

Welchman, A. E., Lam, J. M., & Bulthoff, H. H. (2008). Bayesian motion estimation accounts for a surprising bias in 3d visio. *Proc Natl Acad Sci U S A*, *105*, 12087-12092.

Westheimer, G. (1979). Cooperative neural processes involved in stereoscopic acuity. *Exp Brain Res*, *36*, 585-597.

Westheimer, G., & McKee, S. P. (1978). Steroscopic acuity for moving retinal images. *J Opt Soc Am*, *68*, 450-455.

Westheimer, G., & Mitchell, D. E. (1969). The sensory stimulus for disjunctive eye movements. *Vision Res.*, *9*, 749-755.

Wheatstone, C. (1852). Contributions of the physiology of vision - part the second. on some remarkable, and hitherto unobserved phenomena of binocular vision. *Philosophical Transactions of the Royal Society*, *142*, 1-17.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. fitting, sampling, and goodness of fit. *Percept Psychophys*, *63*, 1293-1313.

Young, M. J., Landy, M. S., & Maloney, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Res.*, *33*, 2685-2696.